Transitional Models For Multivariate Longitudinal Binary Responses With An Application To Behavioral Data Of Canadian Children

A Thesis Submitted to the
College of Graduate Studies and Research
in Partial Fulfillment of the Requirements
for the degree of Master of Science
in the Collaborative Biostatistics Program of
School of Public Health
University of Saskatchewan
Saskatoon

By

Khalid Bin Amin

Permission to Use

In presenting this thesis in partial fulfilment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Director
Collaborative Biostatistics Program
School of Public Health
104 Clinic Place
University of Saskatchewan
Saskatoon, Saskatchewan
Canada
S7N 5E5

Abstract

In longitudinal studies, observational units (commonly referred to as individuals) drawn from some population of interest are followed prospectively over time, and measurements from each individual are taken repeatedly at different points in time with the ultimate goal of characterizing the important features of the population. Longitudinal data naturally arise in many areas of study, where the characterization of the population may be achieved by investigating the effects of covariates on a response. Two or more correlated responses from each individual are also common in longitudinal studies, giving rise to multivariate longitudinal data. For example, the National Longitudinal Survey of Children and Youth (NLSCY) is a long-term study to observe the development of Canadian children. In this survey, measurements about factors influencing a child's social, emotional and behavioral development are collected over time; anxiety and aggression reported for each child in this study may be considered as two response variables to characterize the emotional and behavioral development of children.

Since in longitudinal studies, information is collected repeatedly from each individual over time, the occurrence of an event at a particular time point may increase/decrease the likelihood of the occurrence of another event in future. Failure to take into account this phenomenon in analyzing longitudinal data may lead to erroneous conclusion. Moreover, repeated responses (e.g., anxiety and aggression) from an individual may exhibit correlation over time. Separate analyses of such multivariate longitudinal responses ignore this correlation, and as a result, cannot reveal the potential association among the responses which could be of paramount importance in many applications. Therefore, analysis of multivariate longitudinal data requires substantial extension of the standard longitudinal methods.

In this thesis, we describe a methodology based on the transition models for multivariate longitudinal binary data to address the transitional behavior between two states characterized by binary responses for two different responses (i.e., two processes). Transitional analysis of multivariate longitudinal binary data can address the longitudinal association within processes and enable marginal interpretation of covariate effects. In addition, estimation and inference of the association between the processes can also be achieved via such models. We

illustrate this approach with an application to the NLSCY data, where anxiety and aggression (two correlated responses) are modeled as a function of covariates (gender, depression of person most knowledgeable, number of siblings and family status) to identify their effects on behavioral development of Canadian children. In addition, the extent and direction of the association between two responses are estimated. Gender of the child is found statistically significant for both directions of transition, i.e., from low to high and high to low, of aggression. On contrary, gender of the child is found statistically not significant for both transitions of anxiety. Meanwhile, depression of person most knowledgeable is found marginally significant in the high to low direction for aggression. For association parameters, all four directions of associations between anxiety and aggression are found statistically significant.

Acknowledgements

All praises for the Almighty who accorded me an opportunity for higher study in Canada. Next in order, my most respected personality Dr. Shahedul Ahsan Khan. I would like to express my deepest gratitude and appreciation for his patient supervision, whose invaluable guidance immensely helped the line and the theme of this study, and condescended to enrich the same with all his valuable suggestions. Without his support, both academically and financially, it would not have been possible to complete the degree. I solicit his forgiveness for my time to time intrusions upon his leisure hours for advice. With profound regards, I beg to remain gratefully indebted to him for his labor of kindness throughout this pursuit.

I also convey my profound gratitude to my committee chair and program director Professor Mik Bickis for his wholehearted support during my study in the university. Here I would like to make a special mention of Professor Bickis for his kind and continuous support with necessary funding. I am also thankful to Professor Nazeem Muhajarine, member of my committee, for his valuable comments and guidelines to preparing this thesis.

I am happy to mention and thank the staff, namely, Kyla Duhamel, Marcela Shwydiuk and Margarita Santos of the Department of Mathematics and Statistics, for their sincere support and cooperation. I am grateful to Stephanie Wong and Karen Ruston of School of Public Health for all their support. Besides, the University of Saskatchewan Library deserves a special mention for their help and services.

I would like to thank all my friends in here and back in my country for their support and co-operations during my study. You people were around me all the time to give me courage and mental support.

Also would like to thank my sister, Parveen and her husband, Rahman, and two nieces for all their mental and financial supports throughout my study. Finally, last but not least, would love to express my gratefulness to my parents, Amin and Banu, for giving me all their emotional and financial supports.

Dedicated to my father, Md Ruhul Amin, and mother, Maleka Khurshida Banu.

Contents

List of Tables

List of Figures

List of Abbreviations

APA	American Psychological Association
CD4	Cluster of Differentiation 4
CD8	Cluster of Differentiation 8
HIV	Human Immunodeficiency Virus
NLSCY	National Longitudinal Survey of Children and Youth

Chapter 1

Introduction

In this thesis, we describe a regression model to analyze multivariate longitudinal binary data. The model is based on transitional behavior of individuals over the binary responses which are measured prospectively at fixed points in time during the study period. The model is appealing due to its flexibility and readily interpretable regression coefficients. We illustrate the methodology with an application to the National Longitudinal Survey of Children and Youth (NLSCY) data to infer about the behavioral development of children in Canada.

We start this chapter with an outline of the thesis in Section 1.1. We define the concepts of multivariate longitudinal data in Section 1.2 and present the research problem in Section 1.3. A review of the commonly used statistical techniques to analyze longitudinal binary data is presented in Section 1.4. Then, we introduce the transitional models for longitudinal binary responses in Section 1.5.

1.1 Thesis Overview

In Chapter 2, we describe the transition models (Zeng and Cook, 2007) for multivariate longitudinal binary data, including the mathematical formulation of the model, estimation of the parameters and the asymptotic results for statistical inference. We proceed to Chapter 3 with an overview of the NLSCY data. There the distributions of the covariates and the response variables are presented using frequency distributions and summary statistics. Then, we apply the transition models of Zeng and Cook (2007) to the NLSCY data, and conclude the chapter with scientific conclusion derived from our analysis. We conclude the thesis in Chapter 4 with discussions on finding of this thesis and with some additional considerations, including limitations and future work relevant to this study.

1.2 Multivariate Longitudinal Data

In longitudinal studies, observational units, commonly known as individuals, are drawn from a target population and followed prospectively. Information on variables of interest is collected repeatedly from each individual. Since information is collected repeatedly from each individual, longitudinal study allows us to investigate the within subject changes over time.

Longitudinal data are commonly observed in many research areas, where interest lies in characterizing the important features of specific population (e.g., population of Canadian children) by investigating the effects of covariates (e.g., depression level of person most knowledgeable) on response variable (e.g., anxiety level of children) over time. In addition to single response, two or more correlated responses (e.g., anxiety and aggression) from each individual are also common in longitudinal studies, giving rise to multivariate longitudinal data. For example, the National Longitudinal Survey of Children and Youth (NLSCY) is a long-term study to observe the development of Canadian children over time. Since 1994-1995, a cohort of children is followed progressively and information about various factors influencing a child's social, emotional and behavioral development over time is collected biennially. Information about anxiety and aggression reported for each child in this survey may be considered as two response variables to characterize emotional and behavioral development of children over time.

Since in longitudinal studies, information is collected repeatedly from each individual over time, the occurrence of an event at a particular time point may increase/decrease the likelihood of the occurrence of another event in future. Failure to take into account this phenomenon, commonly known as within-subject variation, in analyzing longitudinal data may lead to erroneous conclusion (Liang and Zeger, 1986). In addition, repeated responses from each individual may exhibit correlation over time (Fitzmaurice, Laird and Ware, 2011; Wu, Liu and Liu, 2009). Separate analyses of such multivariate longitudinal responses ignore this correlation, and as a result, cannot reveal the potential association among the responses which could be of paramount importance in many applications. Therefore, analysis of multivariate longitudinal data requires substantial extension of the standard longitudinal methods.

1.3 The Research Problem

In this section, we present the research problem about the emotional and behavioral development of children over time. We will start with the background of this study in Section 1.3.1, and then we discuss the specific problem to be addressed in this thesis in Section 1.3.2. The objectives of this study is presented in Section 1.3.3.

1.3.1 Background of the Study

In contrast to univariate longitudinal data, multivariate longitudinal data allow us to evaluate the joint association of multiple response variables over time. In many areas of study including biological, environmental, medical and health applications, another important objective involves addressing the question: How changes in covariates over time are associated with two or more responses? Below, we present some examples of how multivariate longitudinal data can arise in different fields of study.

In kidney disease research, one may be interested in studying the effectiveness of biomarkers in identifying the kidney disease and the functionality of kidneys repeatedly (Chapman et. al., 2003), giving rise to two responses from each individual. The same type of scenario is also common in the areas of ophthalmology, audiology and other similar research fields (Zeng and Cook, 2007). In psychiatric studies, researchers are often interested to describe how changes in smoking habit affect the short-term psychometric outcomes (Prochaska and DiClemente, 1983). Also, in comparing the effectiveness of training methods, multiple repeated responses (e.g., behavioral outcomes) from each participant may be viewed as correlated multivariate responses (Cameron et al., 1999). In the field of reliability assessment, interest may lie in identifying how diagnoses of symptoms are assessed by different clinicians at different points in time (Meade et al., 2000; Zeng and Cook, 2007), which leads to multiple outcomes because of different opinions among the clinicians. For HIV patients, interest may lie in joint modeling of CD4 and CD8 counts (Thiebaut et.al., 2002). In all these examples, measurements on multiple outcomes are taken from each individual over time. In the following section, we describe the multivariate longitudinal setup of the research problem of this study.

1.3.2 The National Longitudinal Survey of Children and Youth

Emotional and behavioral development of children may be characterized using their anxiety and aggression levels over time. Identification of the factors associated with the behavioral changes is of great interest among the researches (Dauvergne and Johnson, 2001; Waddell, 2005). Various descriptive analyses were also conducted to study the relationship between anxiety and aggression levels of children, and many studies reported a positive correlation between these two variables (Apter et al., 1990; Barrett et al., 1996; Craig, 1998; Marsee, Weems and Taylor, 2008; Wu, Liu and Liu, 2009).

The National Longitudinal Survey of Children and Youth (NLSCY) is a long-term survey to assess the social, behavioral and emotional development of Canadian children. The survey is a joint venture of Statistics Canada and Human Resources Development Canada (HRDC) (Statistics Canada, 1994). It is a biennial survey – information from each child is collected in every two years of interval. The first cycle (Cycle 1) began in 1994-1995 with a cohort of 22,831 children between the age of 0-11. A child in the survey is expected to remain in the study until he/she reaches the age of 25 (Statistics Canada, 2007). Cycle 8 (2008-2009) is the last known cycle of the NLSCY.

Various types of behavioral and emotional disorders among children and adolescents are recorded in the NLSCY. The person most knowledgeable (PMK) about a child (mostly the mother) was asked a series of questions about the number of times his/her child was involved in various emotional and physical difficulties. Questions related to emotional behavior such as elation, happiness, anxiety and agitation were used to determine the anxiety level of a child. On the other hand, the level of aggression was determined from a child's involvement in and attitude about fighting, threatening and bullying (Statistics Canada, 1994, 1999a; Canadian Institute for Health Information (CIHI), 2009). Factor analysis was carried out to determine scores for each of anxiety and aggression, with a high score indicative of high level of such a disorder (Statistics Canada, 1994). The scores for anxiety range from 0 to 12 for 2-3 year olds children, and from 0 to 14 for 4-11 year olds children. For aggression, the scales are 0-16 and 0-12 for 2-3 and 4-11 year olds children, respectively. For convenience of interpretation, a common practice is to categorize both anxiety and aggression into two levels based on the severity of the problem: high and low. However, there is no concrete guideline to select the cutoff value for dichotomization. According to Statistics Canada (1999, 2007), a youth may be considered to have high level of anxiety/aggression if his/her score is in the highest 10% of the scale. On the other hand, Hotton (2003) suggested to use 80^{th} percentile, a score above this percentile value considered to be high, and low otherwise. In this thesis, we dichotomize anxiety and aggression using the 80^{th} percentile of the respective scores (see Chapter 3 for more detail). Each of the binary responses can be viewed as a two-state process; the status of a child at a given time point can be regarded as a state occupied by the child at that time point for a particular response. In NLSCY, each child is followed over the cycles of the survey, allowing us to observe the transition behavior of a child through the states occupied for both anxiety and aggression over time.

As described above, children participated in NLSCY are followed over time and their anxiety and aggression scores are derived at each cycle. Repeated measurements of characteristics of behavioral development using anxiety and aggression leads to a multivariate longitudinal setup with two responses (anxiety and aggression) observed for each child over time. The behavioral development of children characterized by anxiety and aggression may depend on many factors including gender, depression level of person most knowledgeable, number of siblings and number of hours spent in daycare (Wu, Liu and Liu, 2009). In addition to investigating the inherent association between anxiety and aggression by taking into account their joint effects, identification of the factors significantly associated with the behavioral development of children is also of paramount interest to researchers. Use of an appropriate statistical tool is important to analyze this type of multivariate longitudinal data with the ultimate goal of estimating the parameters of interest that characterize the behavioral development of children over time.

1.3.3 Objectives of the Study

Numerous studies were conducted to investigate the effects of covariates on anxiety and aggression independently, without considering the imminent association between these two response variables (Apter et al., 1990; Archer and Coyne, 2005; Atlas and Pepler, 1998; Bardone et al., 1998; Canadian Paediatric Society, 2009; Craig, 1998; Hotton, 2003). None of these studies addressed the transitional characteristics between the states of these two correlated responses. In this thesis, we consider transition models for multivariate longitudinal binary data to address these gaps. Our ultimate goals are

1. to estimate the marginal transitional probabilities between the states for anxiety and aggression, and to identify the covariates which are significantly associated with these transitions;

2. to estimate the extent and direction of the association between anxiety and aggression of children by taking into account the marginal effects of the covariates on the responses.

In this thesis, we use a dataset which is a subset of the synthetic data files reported by Statistics Canada. This dataset was used for a case study in the 2005 annual conference of the Statistical Society of Canada (Statistical Society of Canada, 2005). It has the first four cycles of the survey: 1994-1995 (Cycle 1), 1996-1997 (Cycle 2), 1998-1999 (Cycle 3) and 2000-2001 (Cycle 4); see Chapter 3 for details. Note that a synthetic data file is a subset of the original data file, and it may include both original and computer-generated observations (Cram, 2004). The main objectives of the synthetic data file is to preserve the identity of the respondents. Although a synthetic data file is a subset of the original data file, outcomes using synthetic data are expected to produce similar results to real data file (Cram, 2004).

1.4 Statistical Models for Longitudinal Binary Data

For longitudinal binary responses, many statistical models have been developed so far, with each model has its own limitations. Among these, the log linear models (Koch et al., 1977) and the probit link models (Ochi and Prentice, 1984) are commonly used to analyze longitudinal binary data. One important drawback of these two methods is that they do not allow time-dependent covariates in the models, whereas it is natural to have time-dependent covariates in a longitudinal study. Stiratelli, Laird and Ware (1984) proposed methods based on the logit link function to overcome this limitation. Logistic growth curve (Korn and Whittemore, 1979), commonly known as logistic-normal model, is also used to model longitudinal binary responses. However, in addition to analytic difficulties and requirement of a large number of repeated observations, this method cannot handle data with low rates of responses (Stiratelli, Laird and Ware, 1984). Recognizing the logistic-normal model, Stiratelli, Laird and Ware (1984) introduced the random-effects models for longitudinal binary response data. Random-effects models (Stiratelli, Laird and Ware, 1984) assume that observations across subjects are independent. Random effects models are useful when the main interest lies in making inference at the individual levels rather than at the population level (Diggle et al., 2002), though interpretation at the population level can be made using these models. Zeger, Liang and Self (1985) proposed first-order Markov chain models to analyze longitudinal binary responses, which can incorporate only time-independent covariates. The likelihood approach of estimation for the random effects and Markov chain models require the joint distribution of outcome variables, the specification of which is not straightforward in practice. Liang and Zeger (1986) introduced the marginal models which do not require the specification of the joint distribution for repeated observations. They developed the estimating equations method (Liang and Zeger, 1986) for statistical inference and showed that the estimators are consistent under some mild conditions (Liang and Zeger, 1986). Binary responses in longitudinal data can also be viewed as a two-state Markov process (Zeng and Cook, 2007). Cox (1972); Zeger, Liang, and Self (1985); and Ware, Lipsitz, and Speizer (1988) discussed the transition models for longitudinal data; where we model the transition probabilities of being in a particular state at a certain time point given the information about the state occupancies at previous time points and a set of covariates. Zeng and Cook (2007) developed a method to generalize the transitional approach for multivariate longitudinal binary data (see Section 1.5 and Chapter 2).

All the methods discussed above can be broadly classified as marginal models (Liang and Zeger, 1986), random-effects models (Stiratelli, Laird and Ware, 1984) and transitional models (Diggle et al., 2002; Fitzmaurice, Laird and Ware, 2011; Zeng and Cook, 2007). In the following subsections, we describe these three approaches in a general context for longitudinal studies.

1.4.1 Marginal Models

In the marginal approach (Liang and Zeger, 1986), the mean response is modeled at each occasion to investigate the overall effects of the covariates, and it does not depend on any random effects and/or previous responses. In contrast to the likelihood approach, the marginal models use the estimating equations method (Liang and Zeger, 1986) to make inference about the regression parameters. One main advantage of the marginal models (Liang and Zeger, 1986) is that it does not require full distribution assumptions of responses, though requires the expressions for the first two moments of responses (Liang and Zeger, 1986; Diggle et al., 2002; Fitzmaurice, Laird and Ware, 2011). Since the marginal models lead to interpretation at population level, such models are also referred as population-averaged model.

1.4.2 Random-Effects Models

The random-effects models (Stiratelli, Laird and Ware, 1984) take into account the potential heterogeneity across individuals (Fitzmaurice et al., 2008; Fitzmaurice, Laird and Ware, 2011) by introducing individual-specific random effects. In such models, it is assumed that individuals have their own trajectories which are assumed to arise from a population of interest (i.e., the individual-specific regression coefficients are assumed random). Since each individual has its own regression coefficients, the random effects models are said to have subject-specific interpretation (Zeng and Cook, 2007).

1.4.3 Transition Models

In transitional modeling approach, the mean response is modeled as a function of the history of the process, which includes information about responses and covariates (both time-independent and time-dependent) at previous time points. One common formulation of the transition models is based on the Markov assumption where the conditional distribution of each response is modeled as an explicit function of previous responses and the covariates (Fitzmaurice et al., 2008). The simplest framework of the transition models for longitudinal data is based on first-order Markov process, where the probability of a transition to a particular state at a future time point depends only on information about the present status and not on the previous states of the process.

1.5 Transition Models for Multivariate Longitudinal Binary Data

In many longitudinal studies with binary responses, simultaneous investigation of covariate effects on transitional behavior of two or more processes over time is of interest. In the case when data on two or more processes are available, an appealing statistical approach is the joint modeling of the responses that allow us not only to investigate the impact of the covariates on the responses, but also to explore the possible dependency between the responses. Under the usual Markov assumption for a single response variable, separate analyses for different response variables may be carried out. However, this approach may lead to erroneous conclusion as it does not take into account the possible dependency among the response variables. In addition, separate analyses cannot reveal the extent and direction of the association between the responses, which is often considered one of the main objectives to the researcher (Zeng and Cook, 2007). Zeng and Cook (2007) developed a method to characterize the transitional behavior for multivariate longitudinal binary responses. Their method of estimation and inference for marginal and association parameters are based on joint analysis of the transition probabilities for multivariate longitudinal binary data. They developed generalized estimating equations (Liang and Zeger, 1986) for statistical inference based on the logistic and alternating logistic models (Carey, Zeger and Diggle, 1993). This approach allows us to identify the covariate effects on marginal transition probabilities in addition to estimate the association parameters for the responses (Zeng and Cook, 2007). In this thesis, we use the multivariate transition models of Zeng and Cook (2007) to characterize the long-term behavioral development of Canadian children. The theoretical and conceptual details of the this method are presented in Chapter 2.

Chapter 2

Transition Models for Multivariate Longitudinal Binary Data

In this chapter, we describe the transitional model of Zeng and Cook (2007) for multivariate longitudinal binary data. As highlighted in Chapter 1, the model is appealing due to its flexibility and great interpretability of regression coefficients. Formulation of the model is presented in Section 2.1. The estimation technique of the regression coefficients is based on the generalized estimating equations (Liang and Zeger, 1986). It involves two stages: (1) estimation of the marginal parameters to investigate the covariate effects, and (2) estimation of the association parameters to assess the relationship between the multiple binary responses. We describe the theoretical details of (1) and (2) in Section 2.2. Asymptotic theory and construction of confidence intervals are presented in Section 2.3. We describe the computational algorithm for estimating regression parameters in Section 2.4, followed by brief description of software and programming language used in the study in Section 2.5. We present derivation of mathematical expressions related to multivariate transition models for longitudinal binary data in Appendices A and B.

2.1 Model Formulation

Suppose there are n individuals and J processes ($J > 1$). For the NLSCY data described in Chapter 1, anxiety and aggression define the two processes with two levels/states (low and high) for each process. Suppose each individual is followed over time, and information is collected about the state occupied by the individual at a particular point in time for process j, ($j = 1, 2, \ldots, J$). For the i^{th} individual ($i = 1, 2, \ldots, n$), let there be K measurements for each process, and let t_k $(k = 1, 2, \ldots, K)$ denote the k^{th} measurement occasion with $t_1 < t_2 < \ldots < t_K$.

Let $Y_i^{(j)}(t_k)$ be the binary response for the state occupied by the i^{th} individual at time t_k for process j, which is defined as follows:

$$Y_i^{(j)}(t_k) = \begin{cases} 1 & \text{if state 2 (high level) is occupied by } i^{th} \text{ individual at time } t_k \text{ for process } j \\ 0 & \text{if state 1 (low level) is occupied by } i^{th} \text{ individual at time } t_k \text{ for process } j \end{cases}.$$

Let there be p_{j-1} covariates associated with process j, including both time-dependent and fixed covariates. Let $\mathbf{x}_i^{(j)}(t_k) = (1, x_{i1}^{(j)}(t_k), x_{i2}^{(j)}(t_k), ..., x_{i,p_j-1}^{(j)}(t_k))'$ be the vector of covariates (including an intercept term) for the i^{th} individual at time t_k for process j. In this study, we consider a common set of covariates for all the processes. Thus, the vector of all covariates at time t_k can be denoted by $\mathbf{x}_i(t_k) = (\mathbf{x}_i^{(1)}(t_k)', \mathbf{x}_i^{(2)}(t_k)', ..., \mathbf{x}_i^{(J)}(t_k)')'$, which is a vector of order $Jp_j \times 1$.

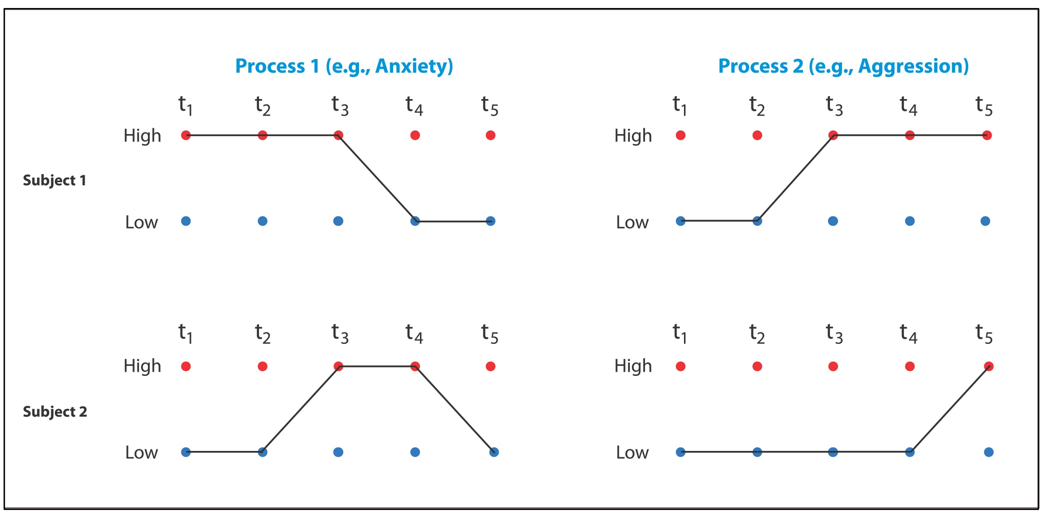

Figure 2.1: A hypothetical display of transitions over time for two individuals with two processes.

Figure (2.1) illustrates the transition path for two individuals with two processes defined by anxiety and aggression. In this figure, subject 1 is at the high level of anxiety at times t_1, t_2 and t_3, experiences a transition to low level at t_4 and then remains in the same state until last time point t_5. Similarly, the same subject is at the low level of aggression at t_1 and

t_2, experiences a transition to the high level at t_3 and occupies this state at t_5. Transition pattern of subject 2 can be described accordingly.

2.1.1 Marginal Models to Estimate the Covariate Effects

In marginal models we investigate the effects of covariates on mean responses at each time point and assume that mean response depends on the covariates of interest, not on any previous responses. Because we mainly focus in modeling mean response at each time point and its dependence on the covariates – marginal models allow us to avoid full distributional assumptions of response vector over time (Fitzmaurice, Laird and Ware, 2011). This is advantageous as joint distribution of discrete longitudinal responses is not easily manageable.

Suppose that the i^{th} individual occupies state l ($l = 1, 2$) at time t_k for process j, and then experiences a transition to state $(3 - l)$ at time t_{k+1}. Under the first-order Markov process, the marginal transition probability of this individual for process j at t_k is defined by

$$\pi_{il}^{(j)}(t_k) = \Pr\left(Y_i^{(j)}(t_{k+1}) = 3 - l | Y_i^{(j)}(t_k) = l,\ \mathbf{x}_i(t_k)\right),\ \text{for } l = 1, 2\cdot \tag{2.1}$$

We assume that probability of transition for an individual from time t_k to t_{k+1} for process j does not depend on the states occupied by this individual for other processes. The marginal transition probabilities can then be modeled using the logistic regression model as

$$\begin{aligned} \pi_{il}^{(j)}(t_k) &= \Pr\left(Y_i^{(j)}(t_{k+1}) = 3 - l | Y_i^{(j)}(t_k) = l,\ \mathbf{x}_i(t_k)\right) \\ &= \frac{e^{\mathbf{x}_i^{(j)}(t_k)'\boldsymbol{\beta}_l^{(j)}}}{1 + e^{\mathbf{x}_i^{(j)}(t_k)'\boldsymbol{\beta}_l^{(j)}}} \end{aligned} \tag{2.2}$$

which can also be expressed using the logit link function as follows:

$$\log\left(\frac{\pi_{il}^{(j)}(t_k)}{1 - \pi_{il}^{(j)}(t_k)}\right) = \text{logit}\left(\pi_{il}^{(j)}(t_k)\right) = \mathbf{x}_i^{(j)}(t_k)'\boldsymbol{\beta}_l^{(j)} \tag{2.3}$$

where $\boldsymbol{\beta}_l^{(j)}$ is a $p_j \times 1$ vector of regression coefficients characterizing the effects of covariates on the transition probabilities out of state l for process j. For a two-state process, the vector of regression coefficients for process j is denoted by $\boldsymbol{\beta}^{(j)} = (\boldsymbol{\beta}_1^{(j)'}, \boldsymbol{\beta}_2^{(j)'})'$, which is a vector of order $2p_j$ with $\boldsymbol{\beta}_1^{(j)'}$ and $\boldsymbol{\beta}_2^{(j)'}$ characterize the effects of the covariates out of state 1 and state 2, respectively. For all J processes, the vector of regression coefficients is denoted by the $(2Jp_j \times 1)$ vector $\boldsymbol{\beta} = (\boldsymbol{\beta}^{(1)'}, \boldsymbol{\beta}^{(2)'}, \cdots, \boldsymbol{\beta}^{(J)'})'$. Note that for $J = 2$, as in the case for the NLSCY data, there are $4p_j$ regression coefficients involved in the marginal models described above.

The marginal probabilities can also be defined using the counting process notation, which is convenient to develop the subsequent theoretical results involving estimation and test of hypothesis. Let $I(A)$ be an indicator function with $I(A) = 1$ if A is true and 0 otherwise. Based on this definition, we introduce two indicator variables as follows: $\delta_{il}^{(j)}(t_k){=}I\Big(Y_i^{(j)}(t_k) = l\Big)$, $l = 1, 2$, indicates whether subject i is at risk of a transition out of state l for process j at time t_k, and $N_{il}^{(j)}(t_k) = I\Big(Y_i^{(j)}(t_k) = l,\ Y_i^{(j)}(t_{k+1}) = 3 - l\Big)$ indicates whether there is any such transition from state l to state $(3 - l)$ for process j. For process j, the vector of the at-risk indicators $\delta_{il}^{(j)}(t_k)$ is expressed as $\boldsymbol{\delta}_i^{(j)}(t_k) = \Big(\delta_{i1}^{(j)}(t_k),\ \delta_{i2}^{(j)}(t_k)\Big)$, $k = 1, 2, \cdots, K$ and $i = 1, 2, \cdots, n$. Then, for all J processes, we define the vector $\boldsymbol{\delta}_i(t_k) = \Big(\boldsymbol{\delta}_i^{(j)}(t_k),\ j = 1, 2, \cdots, J\Big)$.

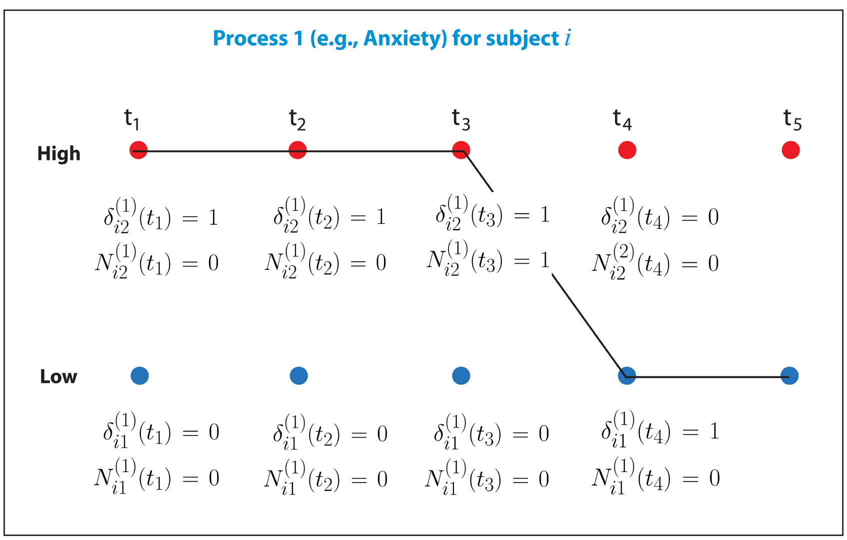

Figure 2.2: A hypothetical presentation of indicator variables $\delta_{il}^{(j)}(t_k)$ and $N_{il}^{(j)}(t_k)$ for the i^{th} individual for a single process e.g., anxiety, over time.

Figure (2.2) illustrates the indicator variables $\delta_{il}^{(j)}(t_k)$ and $N_{il}^{(j)}(t_k)$ based on a hypothetical data for the i^{th} individual for a single process (e.g., anxiety) at five time points. For instance, at times t_1 and t_2 the individual is at high level ($l = 2$) of anxiety, so at risk of transition from this level. We define the indicator variable $\delta_{il}^{(j)}(t_k)$ for time points t_1 and t_2 as $\delta_{i2}^{(1)}(t_1) = \delta_{i2}^{(1)}(t_2) = 1$. For these two time points the individual experiences no transition and the indicator variable $N_{il}^{(j)}(t_k)$ is defined as $N_{i2}^{(1)}(t_1) = N_{i2}^{(1)}(t_2) = 0$. At time point t_3 the individual is still at risk of transition and experiences a transition to low level ($l = 1$) of anxiety. So, the indicator variables at time t_3 are defined as $\delta_{i2}^{(1)}(t_3) = 1$ and $N_{i2}^{(1)}(t_3) = 1$. Similarly, we can define the indicator variables for other time points and levels for the process. Since transition experiences for the last time point t_5 are unknown, we do not define indicator variables for that time point, and consider $5 - 1$ time points in the analysis.

We can now define the transition probabilities using two indicator variables $\delta_{il}^{(j)}(t_k)$ and $N_{il}^{(j)}(t_k)$ as follows:

$$\pi_{il}^{(j)}(t_k) = \Pr\left(N_{il}^{(j)}(t_k) = 1 | \delta_{il}^{(j)}(t_k) = 1, \mathbf{x}_i(t_k)\right) \tag{2.4}$$

which is the probability that the i^{th} individual experiences a transition out of state l at time t_k for process j given the individual's covariate information and that the individual was at risk of a transition at time t_k for the same process, where $i = 1, 2, \cdots, n$, $l = 1, 2$, $k = 1, 2, \cdots, K-1$, and $j = 1, 2, \cdots, J$.

2.1.2 Association Models

The odds ratio is a widely used measure for association between two processes. In general, the odds in favor of an event A is the probability that the event occurs divided by the probability that it does not occur, that is, $\text{ODDS} = \dfrac{\Pr(A)}{1 - \Pr(A)}$. For example, if the probability of a transition is 2/3, then the odds in favor of a transition are $\dfrac{2/3}{1 - 2/3} = 2/1$, that is, it is two times more likely that the transition will occur than it will not occur. The odds ratio (OR) is the ratio of the odds of an event in one group compared to the odds in another group. For example,

$$\text{OR} = \frac{\text{ODDS}\left(N_{\text{low}}^{(\text{anxiety})} | N_{\text{high}}^{(\text{aggression})} = 1\right)}{\text{ODDS}\left(N_{\text{low}}^{(\text{anxiety})} | N_{\text{high}}^{(\text{aggression})} = 0\right)} = 2$$

indicates that a transition out of the low level of anxiety is twice as likely to occur for those who experience a transition out of the high level of aggression than those who did not experience a transition out of the high level of aggression. The association model is developed using the odds ratios of these types (see below). Note that in this thesis, we denote the process of anxiety by j and the process of aggression by j', and the states (i.e., low or high) occupied by an individual for processes j and j' are denoted by l and l', respectively, with $l, l' = 1, 2$ where $1 =$ low and $2 =$ high.

Now, we present the association models for two processes. For correlated binary data, the association in transitions between the two processes can be modeled using the odds ratios (Lipsitz, Laird and Harrington, 1991) for pairs of events $N_{il}^{(j)}(t_k)$ and $N_{il'}^{(j')}(t_k),\ j \neq j'$, where l and l' denote the states for processes j and j', respectively. For the i^{th} individual, the odds of transition out of state l for process j given the individual's transition status for process j' can be defined as

$$\text{ODDS}\left(N_{il}^{(j)}(t_k)|N_{il'}^{(j')}(t_k)\right) = \frac{\Pr\left(N_{il}^{(j)}(t_k) = 1|N_{il'}^{(j')}(t_k), \delta_{il}^{(j)} = 1, \delta_{il'}^{(j')} = 1, \mathbf{x}_i(t_k)\right)}{1 - \Pr\left(N_{il}^{(j)}(t_k) = 1|N_{il'}^{(j')}(t_k), \delta_{il}^{(j)} = 1, \delta_{il'}^{(j')} = 1, \mathbf{x}_i(t_k)\right)}. \tag{2.5}$$

The relevant association parameters between the transitions for the two processes j and j' can be expressed using the odds ratios as follows:

$$\gamma_{i;ll'}^{(jj')}(t_k) = \frac{\text{ODDS}\left(N_{il}^{(j)}(t_k)|N_{il'}^{(j')}(t_k) = 1\right)}{\text{ODDS}\left(N_{il}^{(j)}(t_k)|N_{il'}^{(j')}(t_k) = 0\right)} \tag{2.6}$$

where $j \neq j'$ and $l, l' = 1, 2$. We model these odds ratios using a log-linear model with three covariates, defined using the indicator functions as follows:

$$I(l = 1, l' = 2) = \begin{cases} 1 \text{ if an individual occupies state 1 for process } j \text{ and state 2 for process } j' \\ 0 \text{ otherwise} \end{cases},$$

$$I(l=2,l'=1)=\begin{cases}1 \text{ if an individual occupies state 2 for process } j \text{ and state 1 for process } j' \\ 0 \text{ otherwise}\end{cases},$$

and

$$I(l=2,l'=2)=\begin{cases}1 \text{ if an individual occupies state 2 for process } j \text{ and state 2 for process } j' \\ 0 \text{ otherwise}\end{cases}.$$

Then, the model for the logarithm of the odds ratios $\gamma_{i;ll'}^{(jj')}(t_k)$ at time t_k can be written as

$$\log\left(\gamma_{i;ll'}^{(jj')}(t_k)\right)=\psi_0+\psi_1 I(l=1,l'=2)+\psi_2 I(l=2,l'=1)+\psi_3 I(l=2,l'=2) \quad (2.7)$$

where $\boldsymbol{\psi}=(\psi_0,\psi_1,\psi_2,\psi_3)'$ is the vector of association parameters between two processes j and j'. Note that the model (2.7) reduces to $\log\left(\gamma_{i;11}^{(jj')}(t_k)\right)=\psi_0$ for $l=1$ and $l'=1$, $\log\left(\gamma_{i;22}^{(jj')}(t_k)\right)=\psi_0+\psi_3$ for $l=2$ and $l'=2$, $\log\left(\gamma_{i;12}^{(jj')}(t_k)\right)=\psi_0+\psi_1$ for $l=1$ and $l'=2$ and $\log\left(\gamma_{i;21}^{(jj')}(t_k)\right)=\psi_0+\psi_2$ for $l=2$ and $l'=1$. So, four association parameters for the pairs of transition can be defined for processes j and j': $\gamma_{i;11}^{(jj')}(t_k)=\exp(\psi_0)$ and $\gamma_{i;22}^{(jj')}(t_k)=\exp(\psi_0+\psi_3)$, which are the odds ratios between the two processes for transitions in the same directions (i.e., $l,l'=1$ and $l,l'=2$, respectively), and $\gamma_{i;12}^{(jj')}(t_k)=\exp(\psi_0+\psi_1)$ and $\gamma_{i;21}^{(jj')}(t_k)=\exp(\psi_0+\psi_2)$, which are the odds ratios between the two processes for transitions in the opposite directions (i.e., $l=1,l'=2$ and $l=2,l'=1$, respectively). As an example, $\gamma_{i;11}^{(jj')}(t_k)=\exp(\psi_0)$ indicates how likely it is for a transition out of state 1 for process j among individuals who experience a transition out of state 1 for process j' compared to those who did not experience a transition out of state 1 for process j'.

The odds ratios can also be defined using the conditional joint probabilities for pair of events $(N_{il}^{(j)}(t_k),N_{il'}^{(j')}(t_k))$ at time t_k. The conditional joint probabilities for $(N_{il}^{(j)}(t_k),N_{il'}^{(j')}(t_k))$ at time t_k can be expressed as

$$\pi_{i;ll'}^{(jj')}(t_k)=\Pr\left(N_{il}^{(j)}(t_k)=1,N_{il'}^{(j')}(t_k)=1|\delta_{il}^{(j)}(t_k)=1,\delta_{il'}^{(j')}(t_k)=1,\mathbf{x}_i(t_k)\right), \quad (2.8)$$

which is the probability that an individual experiences transitions at time t_k out of states l and l' for processes j and j', respectively, given the covariates information of the individual and that the individual was at risk of transitions out of states l and l' at time t_k.

Then, the expression of the odds ratios using these conditional joint probabilities can be written as (see Appendix A.1)

$$\gamma_{i;ll'}^{(jj')}(t_k) = \frac{\pi_{i;ll'}^{(jj')}(t_k)\left(1 - \pi_{il}^{(j)}(t_k) - \pi_{il'}^{(j')}(t_k) + \pi_{i;ll'}^{(jj')}(t_k)\right)}{\left(\pi_{il}^{(j)}(t_k) - \pi_{i;ll'}^{(jj')}(t_k)\right)\left(\pi_{il'}^{(j')}(t_k) - \pi_{i;ll'}^{(jj')}(t_k)\right)}. \tag{2.9}$$

Solving (2.9), we can re-express the conditional joint probabilities in terms of the marginal transition probabilities and odds ratios (see Appendix A.2):

$$\pi_{i;ll'}^{(jj')}(t_k) = \begin{cases} \dfrac{a_{i;ll'}^{(jj')}(t_k) - \left(a_{i;ll'}^{(jj')}(t_k)^2 - 4\gamma_{i;ll'}^{(jj')}(t_k)\left(\gamma_{i;ll'}^{(jj')}(t_k) - 1\right) \times \pi_{il}^{(j)}(t_k)\pi_{il'}^{(j')}(t_k)\right)^{1/2}}{2\left(\gamma_{i;ll'}^{(jj')}(t_k) - 1\right)} & \text{if } \gamma_{i;ll'}^{(jj')}(t_k) \neq 1 \\ \pi_{il}^{(j)}(t_k)\pi_{il'}^{(j')}(t_k) \quad \text{otherwise} \end{cases} \tag{2.10}$$

where $a_{i;ll'}^{(jj')}(t_k) = 1 - (1 - \gamma_{i;ll'}^{(jj')}(t_k))(\pi_{il}^{(j)}(t_k) + \pi_{il'}^{(j')}(t_k))$. Thus, conditional joint probabilities can be written in terms of the marginal transition probabilities, $\pi_{il}^{(j)}(t_k)$ and $\pi_{il'}^{(j')}(t_k)$, and the odds ratios $\gamma_{i;ll'}^{(jj')}(t_k)$ which characterize the pairwise transitional association. Note that the definition of the conditional joint probabilities in (2.10) will be used to write down the estimating equations to estimate the model parameters.

2.2 Estimation

In general, the joint distribution of discrete longitudinal responses is relatively complicated. As a result, we can not use the maximum likelihood (ML) methods to estimate the regression parameters to investigate the effects of covariates on binary responses over time. Moreover, the likelihood function assumes independence among responses and ignores possible correlation in response vector. To overcome this problem, Liang and Zeger (1986) introduced the generalized estimating equations (GEE) approach to analyze longitudinal binary responses. Their method of estimating equations does not require the complete specification of joint distribution of responses over time. In addition, GEE (Liang and Zeger, 1986) provides consistent estimates of regression parameters regardless of the true correlation structures between the responses (Fitzmaurice, Laird and Rotnitzky, 1993; Liang and Zeger, 1986). Zeng and Cook (2007) used estimating equations (Liang and Zeger, 1986) approach to estimate marginal transition parameters as well as association parameters. In Section 2.2.1 we present a general idea about the generalized estimating equations (Liang and Zeger, 1986). Then, we discuss the estimating equations (Liang and Zeger, 1986) in context of the multivariate transition models for longitudinal binary response for marginal models in Section 2.2.2, and for association models in Section 2.2.3.

2.2.1 Generalized Estimating Equations (GEE)

Let $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \cdots, Y_{in_i})'$ be a vector of binary response for the i^{th} individual, where $i = 1, 2, \cdots, N$, $j = 1, 2, \cdots, n_i$. Let $\boldsymbol{\mu}_i = (\mu_{i1}, \mu_{i2}, \cdots, \mu_{in_i})$ be the vector of i^{th} individual's mean responses and $\mathbf{x}_i = (\mathbf{x}'_{i1}, \mathbf{x}'_{i2}, \cdots, \mathbf{x}'_{ip_j})'$ be the vector of covariates associated with the individual. Liang and Zeger (1986) introduced the estimating equations methods to analyze longitudinal binary responses. In the generalized estimating approaches we model the marginal mean of response variable, $\mathrm{E}(Y_{ij})$, as a function of covariates for each time point. The marginal models for longitudinal data has the following three specifications as

1. The marginal expectation of response, $\mathrm{E}(Y_{ij}|\mathbf{x}_{ij}) = \mu_{ij}$, depends on covariates, $\mathbf{x}_{ij}$, through a known link function $g(.)$ as

   $$g(\mu_{ij}) = \eta_{ij} = \mathbf{x'}_{ij}\boldsymbol{\beta},$$

   where $\boldsymbol{\beta}$ is a $p_j \times 1$ vector of marginal regression parameters. For longitudinal binary responses, it is common to use the logit link function to model the mean of Y_{ij}, or probability of success, with the covariates as

$$\log\left(\frac{\mu_{ij}}{1-\mu_{ij}}\right) = \text{logit}(\mu_{ij}) = \eta_{ij} = \mathbf{x'}_{ij}\boldsymbol{\beta}, \tag{2.11}$$

2. The variance of each response Y_{ij} for given covariates, $\text{Var}(Y_{ij}|\mathbf{x}_{ij})$, depends on marginal mean μ_{ij} according to

   $$\text{Var}(Y_{ij}|\mathbf{x}_{ij}) = \phi v(\mu_{ij}),$$

   where $v(\mu_{ij})$ is some known function of marginal means, and ϕ is a scale parameter that may be known or may need to be estimated. For binary responses, variance of responses is defined as $\text{Var}(Y_{ij}|\mathbf{x}_{ij}) = \mu_{ij}(1-\mu_{ij})$ and ϕ is fixed at 1.

3. The pairwise within-subject association between repeated responses Y_{ij} and Y_{ik} for given covariates are assumed to be a function of marginal means and an additional within-subject association parameters, $\boldsymbol{\alpha}$, i.e., $\text{Corr}(Y_{ij}, Y_{ik}) = \rho(\mu_{ij}, \mu_{ik}; \boldsymbol{\alpha})$.

The generalized estimating equations for $\boldsymbol{\beta}$ are defined as

$$\mathbf{U}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \mathbf{D'}_i \mathbf{V}_i^{-1}(\mathbf{Y}_i - \boldsymbol{\mu}_i) = \mathbf{0} \tag{2.12}$$

where $\mathbf{D}_i = \dfrac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta'}}$ is the partial derivative matrix of mean responses with respect to marginal parameters $\boldsymbol{\beta}$ and $\mathbf{V}_i$ is the approximation to true covariance matrix of responses, commonly known as working covariance matrix.

The corresponding working covariance matrix in (2.12) can be defined as the product of standard deviations and correlations between pairwise responses as

$$\mathbf{V}_i = \mathbf{A}_i^{1/2}[\text{Corr}(\mathbf{Y}_i)]\mathbf{A}_i^{1/2} \tag{2.13}$$

where $\mathbf{A}_i$ is a diagonal matrix, i.e., $\mathbf{A}_i = \text{diag}\,[\text{Var}(Y_{i1}), \text{Var}(Y_{i2}), \cdots, \text{Var}(Y_{in_i})]$, with elements $\text{Var}(Y_{ij}|\mathbf{x}_{ij}) = \phi v(\mu_{ij})$ and $\text{Corr}(\mathbf{Y}_i)$ is the working correlation matrix, here a function of association parameters $\boldsymbol{\alpha}$.

The generalized estimating approach provides consistent estimators of regression parameters given that the model for mean responses is specified correctly to the vector of responses (Fitzmaurice, Laird and Rotnitzky, 1993). Since the estimating equations depend on both regression parameters $\boldsymbol{\beta}$ and association parameter $\boldsymbol{\alpha}$, a two stage iterative method is required for estimation.

2.2.2 Estimating Equations for the Marginal Models

The marginal interpretation of the covariate effects are based on the estimates of the regression coefficients $\boldsymbol{\beta}$ (see Section 2.1.1). For example, $\exp\left(\hat{\beta}^{(j)}_{1,\text{gender}}\right) = 1.9$ indicates that males are almost twice as likely to experience a transition out of state 1 (i.e., low $\rightarrow$ high) than females for process j. In this section, we describe the GEE approach to estimate $\boldsymbol{\beta}$.

For given covariates $\mathbf{x}_i(t_k)$, the marginal distribution of the processes can be characterized using the first-order Markov chain. Then, for n independent individuals, the likelihood function for process j can be expressed in the form of product binomials as follows:

$$\prod_{i=1}^{n}\prod_{k=1}^{K-1}\prod_{l=1}^{2}\left(\left(\pi^{(j)}_{il}(t_k)\right)^{N^{(j)}_{il}(t_k)} \times \left(1-\pi^{(j)}_{il}(t_k)\right)^{1-N^{(j)}_{il}(t_k)}\right)^{\delta^{(j)}_{il}(t_k)}. \tag{2.14}$$

The log-likelihood function for process j can then be written as

$$\ell_j\left(\boldsymbol{\beta}^{(j)}\right) = \sum_{i=1}^{n}\sum_{k=1}^{K-1}\sum_{l=1}^{2}\delta^{(j)}_{il}(t_k)\left\{N^{(j)}_{il}(t_k)\log\left(\pi^{(j)}_{il}(t_k)\right) + \left(1-N^{(j)}_{il}(t_k)\right)\log\left(1-\pi^{(j)}_{il}(t_k)\right)\right\} \tag{2.15}$$

for all $j = 1, 2, ..., J$.

To derive the estimating equations for the parameter vector $\boldsymbol{\beta}$ under the first-order Markov process, we define an indicator variable

$$N^{(j)}_{i}(t_k) = \sum_{l=1,2}\delta^{(j)}_{il}(t_k)N^{(j)}_{il}(t_k) \tag{2.16}$$

where $N^{(j)}_{i}(t_k)$ indicates whether there is any transition out of either state $l = 1, 2$ for subject i at time t_k for the process j. The vector of these transition indicators is denoted by

$\mathbf{N}_i(t_k) = (N_i^{(j)}(t_k); j = 1, 2, \cdots, J)', \ i = 1, 2, \cdots, n$ and $k = 1, 2, \cdots, K-1$.

For given $\boldsymbol{\delta}_i(t_k) = (\boldsymbol{\delta}_i^{(j)}(t_k), j = 1, 2, \cdots, J)$ and $\mathbf{x}_i(t_k)$, the conditional expectation of $N_i^{(j)}(t_k)$ becomes (see Appendix A.3)

$$\mu_i^{(j)}(t_k) = \mathrm{E}\left(N_i^{(j)}(t_k)|\boldsymbol{\delta}_i(t_k), \mathbf{x}_i(t_k)\right) = \sum_{l=1,2} \delta_{il}^{(j)}(t_k)\pi_{il}^{(j)}(t_k) \ \cdot \tag{2.17}$$

We denote the vector of all conditional expectations by $\boldsymbol{\mu}_i(t_k) = (\mu_i^{(j)}(t_k), \ j = 1, 2, ..., J)'$.

The estimation of $\boldsymbol{\beta}$ can be based on the generalized estimating equations (Zeng and Cook, 2007), defined by

$$\mathbf{U}_1(\boldsymbol{\beta}, \boldsymbol{\psi}) = \sum_{i=1}^{n}\sum_{k=1}^{K-1} D_i(t_k; \boldsymbol{\beta})' V_i(t_k; \boldsymbol{\beta}, \boldsymbol{\psi})^{-1}\mathbf{R}_i(t_k) = 0 \tag{2.18}$$

where $\mathbf{R}_i(t_k) = (R_i^{(1)}(t_k), R_i^{(2)}(t_k), ..., R_i^{(J)}(t_k))'$ with $R_i^{(j)}(t_k) = N_i^{(j)}(t_k) - \mu_i^{(j)}(t_k)$ being the error terms; $D_i(t_k; \boldsymbol{\beta}) = \dfrac{\partial \boldsymbol{\mu}_i(t_k)}{\partial \boldsymbol{\beta}'}$ is the $J \times (2\sum_{j=1}^{J} p_j)$ matrix of partial derivatives of the conditional expectations with respect to $\boldsymbol{\beta}$; and $V_i(t_k; \boldsymbol{\beta}, \boldsymbol{\psi})$ is the $J \times J$ working covariance matrix of $\mathbf{N}_i(t_k)$ conditional on $\boldsymbol{\delta}_i(t_k)$ and $\mathbf{x}_i(t_k)$ (see Appendix A.4). The (j, j') entry of $V_i(t_k; \boldsymbol{\beta}, \boldsymbol{\psi})$ can be expressed as (see Appendix A.5)

$$v_i^{(jj')}(t_k) = \mathrm{Cov}\left(N_i^{(j)}(t_k), N_i^{(j')}(t_k)|\boldsymbol{\delta}_i(t_k), \mathbf{x}_i(t_k)\right), \tag{2.19}$$

where

$$v_i^{(jj')}(t_k) = \begin{cases} \sum_{l=1,2} \delta_{il}^{(j)}(t_k)\pi_{il}^{(j)}(t_k)\left(1 - \pi_{il}^{(j)}(t_k)\right), & \text{if } j = j' \\ \sum_{l,l'=1,2} \delta_{il}^{(j)}(t_k)\left(\pi_{i;ll'}^{(jj')}(t_k) - \pi_{il}^{(j)}(t_k)\pi_{il'}^{(j')}(t_k)\right), & \text{if } j \neq j' \end{cases} \ . \tag{2.20}$$

Note that (2.18) depends on $\gamma_{i;ll'}^{(jj')}(t_k)$, and hence on $\boldsymbol{\psi}$ via $\pi_{i;ll'}^{(jj')}(t_k)$. So, given $\gamma_{i;ll'}^{(jj')}(t_k)$ or $\boldsymbol{\psi}$, we can solve (2.18) numerically to estimate $\boldsymbol{\beta}$. The left side of (2.18) is an unbiased estimating function, and so the solution is consistent for $\boldsymbol{\beta}$ (Zeng and Cook, 2007). We can show this by taking expectation of $R_i^{(j)}(t_k)$ with respect to the true distribution as follows:

$$
\begin{aligned}
\mathrm{E}\left(R_i^{(j)}(t_k)|\mathbf{x}_i(t_k)\right) &= \mathrm{E}\left(N_i^{(j)}(t_k) - \mu_i^{(j)}(t_k)|\mathbf{x}_i(t_k)\right) \\
&= \mathrm{E}_{\boldsymbol{\delta}_i(t_k)|\mathbf{x}_i(t_k)}\left(\mathrm{E}_{N_i^{(j)}(t_k)|\boldsymbol{\delta}_i(t_k),\mathbf{x}_i(t_k)}\left(N_i^{(j)}(t_k) - \mu_i^{(j)}(t_k)\right)\right).
\end{aligned} \tag{2.21}
$$

Since $\mathrm{E}(N_i^{(j)}(t_k)) = \mu_i^{(j)}(t_k)$ by (2.17), the above expectation is zero (see Zeng and Cook (2007) for detail), and therefore the left side of (2.18) is an unbiased estimation function.

2.2.3 Estimating Equations for the Association Models

Let $\xi_i^{(jj')}(t_k)$ be the conditional expectation of $N_i^{(j)}(t_k)$ for process j given $N_i^{(j')}(t_k)$ for process j'. This conditional expectation can be written as (see Appendix B.1)

$$
\begin{aligned}
\xi_i^{(jj')}(t_k) &= \mathrm{E}\left(N_i^{(j)}(t_k)|N_i^{(j')}(t_k), \boldsymbol{\delta}_i(t_k), \mathbf{x}_i(t_k)\right) \\
&= \sum_{l,l'=1,2} \delta_{il}^{(j)}(t_k)\delta_{il'}^{(j')}(t_k)\xi_{i;ll'}^{(jj')}(t_k)
\end{aligned} \tag{2.22}
$$

where $\xi_{i;ll'}^{(jj')}(t_k) = \Pr\left(N_{il}^{(j)}(t_k) = 1|N_{il'}^{(j')}(t_k), \delta_{il}^{(j)}(t_k) = 1, \delta_{il'}^{(j')}(t_k) = 1, \mathbf{x}_i(t_k)\right)$, for $j < j'$. It can also be expressed in terms of the transition probabilities and the odds ratios (Carey, Zeger and Diggle, 1993; Diggle, 1992; see also Appendix B.2) as follows:

$$
\operatorname{logit}\left(\xi_{i;ll'}^{(jj')}(t_k)\right) = \log\left(\gamma_{i;ll'}^{(jj')}(t_k)\right) \times N_i^{(j')}(t_k) + \log\left(\frac{\pi_{il}^{(j)}(t_k) - \pi_{i;ll'}^{(jj')}(t_k)}{1 - \pi_{il}^{(j)}(t_k) - \pi_{il'}^{(j')}(t_k) + \pi_{i;ll'}^{(jj')}(t_k)}\right) \tag{2.23}
$$

which is equivalent to

$$
\xi_{i;ll'}^{(jj')}(t_k) = \frac{\mathrm{P}_{i;ll'}^{(jj')}(t_k)\left(\gamma_{i;ll'}^{(jj')}(t_k)\right)^{N_i^{(j')}(t_k)}}{1 + \mathrm{P}_{i;ll'}^{(jj')}(t_k)\left(\gamma_{i;ll'}^{(jj')}(t_k)\right)^{N_i^{(j')}(t_k)}} \tag{2.24}
$$

where, $\mathrm{P}_{i;ll'}^{(jj')}(t_k) = \dfrac{\pi_{il}^{(j)}(t_k) - \pi_{i;ll'}^{(jj')}(t_k)}{1 - \pi_{il}^{(j)}(t_k) - \pi_{il'}^{(j')}(t_k) + \pi_{i;ll'}^{(jj')}(t_k)}$.

Now, let $\epsilon_i^{(jj')}(t_k) = N_i^{(j)}(t_k) - \xi_i^{(jj')}(t_k)$ be the conditional residual for the i^{th} individual at time t_k for process $j < j'$. The vector of these conditional residuals is denoted by $\boldsymbol{\epsilon}_i(t_k) = (\epsilon_i^{(jj')}(t_k),\ j < j')$. Based on the conditional expectations in (2.22) and the conditional residuals, we can formulate the estimating equations for the association parameters $\boldsymbol{\psi}$ as follows (Zeng and Cook, 2007):

$$\mathbf{U}_2(\boldsymbol{\beta}, \boldsymbol{\psi}) = \sum_{i=1}^{n} \sum_{k=1}^{K-1} C_i(t_k|\boldsymbol{\beta}, \boldsymbol{\psi})' S_i(t_k|\boldsymbol{\beta}, \boldsymbol{\psi})^{-1} \boldsymbol{\epsilon}_i(t_k) = 0 \tag{2.25}$$

where $C_i(t_k|\boldsymbol{\beta}, \boldsymbol{\psi}) = \dfrac{\partial \boldsymbol{\xi}_i(t_k)}{\partial \boldsymbol{\psi}'}$ (see Appendix B.3) is the matrix of partial derivatives of $\boldsymbol{\xi}_i(t_k) = \left(\xi_i^{(jj')}(t_k),\ j < j'\right)'$ with respect to $\boldsymbol{\psi}$, and $S_i(t_k|\boldsymbol{\beta}, \boldsymbol{\psi}) = \text{diag}\left\{\xi_i^{(jj')}(t_k)\left(1 - \xi_i^{(jj')}(t_k)\right), j < j'\right\}$ is the working covariance matrix of $\boldsymbol{\epsilon}_i(t_k)$, which are assumed independent.

Note that (2.25) depends on $\xi_{i;ll'}^{(jj')}(t_k)$, and hence on $\gamma_{i;ll'}^{(jj')}(t_k)$ and $\pi_{i;ll'}^{(jj')}(t_k)$. So, for given $\boldsymbol{\beta}$ we can solve (2.25) numerically to estimate $\boldsymbol{\psi}$.

In summary, estimation of the parameters are based on the generalized estimating equations (2.18) and (2.25). It can be shown that the estimators of the marginal transition parameters $\widehat{\boldsymbol{\beta}}$ and the association parameters $\psi$are consistent estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\psi}$, given that models for both transition probabilities and odds ratios are specified correctly.

2.3 Asymptotic Theory and Confidence Intervals

Let $\boldsymbol{\zeta} = (\boldsymbol{\beta}', \boldsymbol{\psi}')'$ be the vector of all the parameters to be estimated by solving (2.18) and (2.25) iteratively using the Newton-Raphson algorithm. Partial derivatives of estimating equations (2.18) and (2.25) with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\psi}$ is given as

$$\begin{aligned} M_1(\boldsymbol{\beta}, \boldsymbol{\psi}) &= \frac{\partial \mathbf{U}_1(\boldsymbol{\beta}, \boldsymbol{\psi})}{\partial \boldsymbol{\beta}} \\ &= \sum_{i=1}^{n} \sum_{k=1}^{K-1} D_i(t_k; \boldsymbol{\beta})' V_i(t_k; \boldsymbol{\beta}, \boldsymbol{\psi})^{-1} D_i(t_k; \boldsymbol{\beta}) \end{aligned} \tag{2.26}$$

and

$$M_2(\boldsymbol{\beta},\boldsymbol{\psi}) = \frac{\partial \mathbf{U}_2(\boldsymbol{\beta},\boldsymbol{\psi})}{\partial \boldsymbol{\psi}}$$
$$= \sum_{i=1}^{n}\sum_{k=1}^{K-1} C_i(t_k;\boldsymbol{\beta})' S_i(t_k;\boldsymbol{\beta},\boldsymbol{\psi})^{-1} C_i(t_k;\boldsymbol{\beta}). \quad (2.27)$$

For given values of $\widehat{\boldsymbol{\beta}}_{(r)}$ and $\widehat{\boldsymbol{\psi}}_{(r)}$, solutions for (2.18) and (2.25) are obtained by using the iterative method. We can update the values of $\widehat{\boldsymbol{\beta}}_{(r)}$ and $\widehat{\boldsymbol{\psi}}_{(r)}$ using the following two equations as

$$\widehat{\boldsymbol{\beta}}_{(r+1)} = \widehat{\boldsymbol{\beta}}_{(r)} - \frac{\mathbf{U}_1\left(\widehat{\boldsymbol{\beta}}_{(r)}, \widehat{\boldsymbol{\psi}}_{(r)}\right)}{M_1\left(\widehat{\boldsymbol{\beta}}_{(r)}, \widehat{\boldsymbol{\psi}}_{(r)}\right)} \quad (2.28)$$

and

$$\widehat{\boldsymbol{\psi}}_{(r+1)} = \widehat{\boldsymbol{\psi}}_{(r)} - \frac{\mathbf{U}_2\left(\widehat{\boldsymbol{\beta}}_{(r)}, \widehat{\boldsymbol{\psi}}_{(r)}\right)}{M_2\left(\widehat{\boldsymbol{\beta}}_{(r)}, \widehat{\boldsymbol{\psi}}_{(r)}\right)}. \quad (2.29)$$

After each successful iteration, we continue to update the values of regression parameters $\boldsymbol{\beta}$ and $\boldsymbol{\psi}$ iteratively using (2.28) and (2.29). We continue this process until $\widehat{\boldsymbol{\zeta}}_{(r)} = (\widehat{\boldsymbol{\beta}}'_{(r)}, \widehat{\boldsymbol{\psi}}'_{(r)})'$ converges to $\widehat{\boldsymbol{\zeta}} = (\widehat{\boldsymbol{\beta}}', \widehat{\boldsymbol{\psi}}')'$. We describe the computational algorithm of this iterative process in Section 2.4. Zeng and Cook (2007) shown that the resulting estimates of $\widehat{\boldsymbol{\zeta}} = (\widehat{\boldsymbol{\beta}}', \widehat{\boldsymbol{\psi}}')'$ are consistent and asymptotically follow multivariate normal distribution. To see that, let us define the estimating equations of marginal and association parameters for n individuals over K time points as

$$\mathbf{U}(\boldsymbol{\zeta}) = \sum_{i=1}^{n}\sum_{k=1}^{K-1} \mathbf{U}_{ik}(\boldsymbol{\zeta})$$

$$= \begin{pmatrix} \sum_{i=1}^{n}\sum_{k=1}^{K-1} D_i(t_k;\boldsymbol{\beta},\boldsymbol{\psi})' V_i(t_k;\boldsymbol{\beta},\boldsymbol{\psi})^{-1} \boldsymbol{R}_i(t_k) \\ \sum_{i=1}^{n}\sum_{k=1}^{K-1} C_i(t_k;\boldsymbol{\beta},\boldsymbol{\psi})' S_i(t_k;\boldsymbol{\beta},\boldsymbol{\psi})^{-1} \boldsymbol{\epsilon}_i(t_k) \end{pmatrix}. \quad (2.30)$$

Using the first-order Taylor series expansion we can find the asymptotic distribution of the estimated parameters. The first-order Taylor series of a function $f(x)$ at x_0 is defined as

$$f(x_r) = f(x_0) + f'(x_0)(x_r - x_0) \tag{2.31}$$

where x_r is the root of the function $f(x)$ and x_0 is an estimate of x_r. Let $f'(x_0)$ be the first order derivative of the function at x_0. The error term can be obtained from (2.31) as

$$(x_r - x_0) = -\frac{f(x_0)}{f'(x_0)}. \tag{2.32}$$

From (2.32), we can define the first-order Taylor series expansion of $\mathbf{U}(\widehat{\boldsymbol{\zeta}})$ about $\boldsymbol{\zeta}$. The error between estimated regression parameters $\widehat{\boldsymbol{\zeta}}$ and $\boldsymbol{\zeta}$ can be derived as

$$\left(\widehat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}\right) = -\frac{\displaystyle\sum_{i=1}^{n}\sum_{k=1}^{K-1}\mathbf{U}_{ik}(\boldsymbol{\zeta})}{\displaystyle\sum_{i=1}^{n}\sum_{k=1}^{K-1}\frac{\partial\mathbf{U}_{ik}(\boldsymbol{\zeta})}{\partial(\boldsymbol{\zeta})}}. \tag{2.33}$$

Multiplying both sides of (2.33) by $\sqrt{n(K-1)}$, we can re-write the equation as

$$\sqrt{n(K-1)}\left(\widehat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}\right) = -\frac{\displaystyle\frac{1}{\sqrt{n(K-1)}}\sum_{i=1}^{n}\sum_{k=1}^{K-1}\mathbf{U}_{ik}(\boldsymbol{\zeta})}{\displaystyle\frac{1}{n(K-1)}\sum_{i=1}^{n}\sum_{k=1}^{K-1}\frac{\partial\mathbf{U}_{ik}(\boldsymbol{\zeta})}{\partial(\boldsymbol{\zeta})}}. \tag{2.34}$$

If the models for the marginal transitional probabilities and association between transitions are specified correctly, $\mathrm{E}(\mathbf{U}_{ik}(\boldsymbol{\zeta})) = 0$ (Zeng and Cook, 2007).Using the central limit theorem (CLT), it can be shown that $\displaystyle\frac{1}{\sqrt{n(K-1)}}\sum_{i=1}^{n}\sum_{k=1}^{K-1}\mathbf{U}_{ik}(\boldsymbol{\zeta})$ converge to multivariate normal (MVN) distribution with mean $\mathbf{0}$ and variance $I(\boldsymbol{\zeta})$, i.e.,

$$\frac{1}{\sqrt{n(K-1)}}\sum_{i=1}^{n}\sum_{k=1}^{K-1}\mathbf{U}_{ik}(\boldsymbol{\zeta}) \xrightarrow{d} \mathrm{MVN}(\mathbf{0}, I(\boldsymbol{\zeta})), \tag{2.35}$$

where $I(\boldsymbol{\zeta}) = \mathrm{E}(\mathbf{U}_{ik}(\boldsymbol{\zeta})(\mathbf{U}_{ik}{}'(\boldsymbol{\zeta}))$. Also, let us define the expected value of $\dfrac{\partial \mathbf{U}_{ik}(\boldsymbol{\zeta})}{\partial \boldsymbol{\zeta}}$ as $\Gamma(\boldsymbol{\zeta}) = \mathrm{E}\left(\dfrac{\partial \mathbf{U}_{ik}(\boldsymbol{\zeta})}{\partial \boldsymbol{\zeta}}\right)$.

According to the law of large numbers, it can be shown that $\dfrac{1}{n(K-1)} \displaystyle\sum_{i=1}^{n} \sum_{k=1}^{K-1} \dfrac{\partial \mathbf{U}_{ik}(\boldsymbol{\zeta})}{\partial (\boldsymbol{\zeta})}$ converge to expected values $\Gamma(\boldsymbol{\zeta})$ with probability, i.e.,

$$\frac{1}{n(K-1)} \sum_{i=1}^{n} \sum_{k=1}^{K-1} \frac{\partial \mathbf{U}_{ik}(\boldsymbol{\zeta})}{\partial (\boldsymbol{\zeta})} \xrightarrow{P} \Gamma(\boldsymbol{\zeta}). \tag{2.36}$$

Since the numerator and denominator of (2.34) converges to distribution and probability respectively, following the Slutzky's theorem it can be shown that error of estimated regression parameters converge to multivariate normal distribution with mean $\mathbf{0}$ and variance $\Gamma^{-1}(\boldsymbol{\zeta})I(\boldsymbol{\zeta})[\Gamma^{-1}(\boldsymbol{\zeta})]'$. Thus,

$$\sqrt{n(K-1)}(\widehat{\boldsymbol{\zeta}} - \boldsymbol{\zeta}) \xrightarrow{d} \mathrm{MVN}(\mathbf{0}, \Gamma^{-1}(\boldsymbol{\zeta})I(\boldsymbol{\zeta})[\Gamma^{-1}(\boldsymbol{\zeta})]')\cdot \tag{2.37}$$

The asymptotic variance of $\boldsymbol{\zeta}$ can be consistently estimated as

$$\widehat{V}_{ALR} = \frac{1}{n(K-1)} \widehat{\Gamma}^{-1}(\widehat{\boldsymbol{\zeta}})\widehat{I}(\widehat{\boldsymbol{\zeta}})[\widehat{\Gamma}^{-1}(\widehat{\boldsymbol{\zeta}})]' \tag{2.38}$$

with $\widehat{I}(\widehat{\boldsymbol{\zeta}}) = \dfrac{\sum_{i=1}^{n} \sum_{i=1}^{K-1} \mathbf{U}_{ik}(\widehat{\boldsymbol{\zeta}})\mathbf{U}'_{ik}(\widehat{\boldsymbol{\zeta}})}{n(K-1)}$

and

$$\widehat{\Gamma}(\widehat{\boldsymbol{\zeta}}) = \frac{1}{n(K-1)} \begin{pmatrix} M_1(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\psi}}) & \mathbf{0} \\ M_{21}(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\psi}}) & M_2(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\psi}}) \end{pmatrix}, \tag{2.39}$$

where M_1 and M_2 can be derived using (2.26) and (2.27) respectively, and $M_{21}(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\psi}}) = \sum_{i=1}^{n} \sum_{k=1}^{K-1} C_i(t_k; \boldsymbol{\beta})' S_i(t_k; \boldsymbol{\beta}, \boldsymbol{\psi})^{-1} \dfrac{\partial \boldsymbol{\xi}(t_k)}{\partial \boldsymbol{\beta}'}$.

A $100(1-\alpha)\%$ confidence interval for the marginal transition coefficients $\boldsymbol{\beta}$ and the association parameters $\boldsymbol{\psi}$ are computed using $\widehat{\boldsymbol{\zeta}} \pm Z_{\alpha/2}\mathrm{SE}(\widehat{\boldsymbol{\zeta}})$ where $\mathrm{SE}(\widehat{\boldsymbol{\zeta}})$ can be obtained using (2.38). We can also calculate the $100(1-\alpha)\%$ confidence intervals for odds ratios by exponentiating the confidence limits for $\boldsymbol{\zeta}$: $\exp\left[\widehat{\boldsymbol{\zeta}} \pm Z_{\alpha/2}\mathrm{SE}(\widehat{\boldsymbol{\zeta}})\right]$.

2.4 Computational Algorithm

The estimating equations for the marginal regression coefficients $\boldsymbol{\beta}$ is $\mathbf{U}_1$ (2.18) and those for the association parameters $\boldsymbol{\psi}$ (or $\boldsymbol{\gamma}$) is $\mathbf{U}_2$ (2.25). We can solve these two equations iteratively to estimate $\boldsymbol{\beta}$ and $\boldsymbol{\psi}$. Recall that $\boldsymbol{\zeta} = (\boldsymbol{\beta}', \boldsymbol{\psi}')'$ is the vector of all the parameters collectively. Then, $\boldsymbol{\beta}$ and $\boldsymbol{\psi}$ can be estimated by solving $\mathbf{U}_1$ and $\mathbf{U}_2$ iteratively using the Newton-Raphson algorithm as described in Section 2.3. We present the steps of this iterative process as follows:

Step 1: Let $\boldsymbol{\beta}^{(0)}$ and $\boldsymbol{\psi}^{(0)}$ be the initial estimates. The estimating equations $\mathbf{U}_1$ are used to update $\boldsymbol{\beta}^{(0)}$. We denote this update by $\boldsymbol{\beta}^{(1)}$. Given $\boldsymbol{\beta}^{(1)}$ and $\boldsymbol{\psi}^{(0)}$, we use $\mathbf{U_2}$ to update the association parameter vector, denoted by $\boldsymbol{\psi}^{(1)}$. The updated estimates from step 1 are then $(\boldsymbol{\beta}^{(1)}, \boldsymbol{\psi}^{(1)})$.

Step 2: Given $(\boldsymbol{\beta}^{(1)}, \boldsymbol{\psi}^{(1)})$ we update $\boldsymbol{\beta}^{(1)}$ using $\mathbf{U}_1$, and we denote the update by $\boldsymbol{\beta}^{(2)}$. Then given $\boldsymbol{\beta}^{(2)}$ and $\boldsymbol{\psi}^{(1)}$ we update $\boldsymbol{\psi}^{(1)}$ using $\mathbf{U}_2$, and denote it by $\boldsymbol{\psi}^{(2)}$. Thus, the updated estimates from step 2 are $(\boldsymbol{\beta}^{(2)}, \boldsymbol{\psi}^{(2)})$.

We repeat the above two steps until convergence. We consider $1\mathrm{e}^{-05}$ as the tolerance limit. Letting $\boldsymbol{\zeta}_{(r)}$ and $\boldsymbol{\zeta}_{(r+1)}$ to be the updates from iterations r and $r+1$, respectively, the above process is continued until $\max(|\boldsymbol{\zeta}_{(r)} - \boldsymbol{\zeta}_{(r+1)}|)$ is less than the tolerance limit. Then, $\boldsymbol{\zeta}_{(r+1)}$ is taken as the final estimate of the parameters.

2.5 Software Implementation

We used the statistical software R (R Core Team, 2013) to produce the frequency distributions and the summary statistics (see Chapter 3) to describe the distributions of the covariates.

We also fit the models using R. We wrote R code to implement the algorithm described above. To solve the generalized estimating equations (2.18) and (2.25) at each step of the algorithm using the Newton-Raphson method, we used the R function `nleqslv` (Hasselman, 2013), which was written to solve system of nonlinear equations. The R code used in this thesis is presented in Appendix C.

Chapter 3

Data and Results

In this chapter, we first present an overview of the children behavior data, and then present our findings about the marginal effects of the covariates on anxiety and aggression along with the estimates of the association parameters between these two responses.

We begin this chapter with a brief description of the National Longitudinal Survey of Children and Youth in section 3.1. In Section 3.2, we present the structure and source of the data. Then we describe the variables under study in Section 3.3. The analysis of the data with interpretation of the parameters are presented in Section 3.4. A discussion about our findings is presented in Section 3.5.

3.1 National Longitudinal Survey of Children and Youth

The National Longitudinal Survey of Children and Youth (NLSCY) is a long-term longitudinal study to monitor the development and well-being of Canadian children from their birth to early adulthood. They survey is jointly conducted by Statistics Canada and Human Resources Development Canada (HRDC), (Statistics Canada, 1994) presently known as Human Resources and Skills Development Canada (HRSDC). In addition to different demographic characteristics, information about factors influencing social, emotional as well as behavioral development are collected repeatedly, allowing to monitor the effects of these factors on the child's development longitudinally. Although the initial strategy was to create a cohort of children aged 0 to 11 years and to follow them in every two-year intervals until they reached the age of 25, the survey is continued to include more children as it progresses over time, leading to more information about the development of children in Canada.

Cycle 1 of the survey was carried out in 1994-1995 which consisted of a sample of 22,813 children from new born to 11 years of age. In most cases, information in the survey was collected from parents; the person most knowledgeable (PMK) about a child, usually the mother, was interviewed to collect information on social, behavioral, health and demographic characteristics of the child. The last known cycle of the survey was Cycle 8 which took place in 2008-2009.

3.2 Data Source

In this thesis, we consider a dataset reported by the Statistical Society of Canada (SSC) in its 2005 annual meeting (Statistical Society of Canada, 2005). The dataset was used for a case study competition for students. It includes four cycles of the NLSCY data, conducted in 1994-1995 (Cycle 1), 1996-1997 (Cycle 2), 1998-1999 (Cycle 3) and 2000-2001 (Cycle 4). The reported dataset is a sample of the synthetic datafile from the NLSCY, which was selected based on the following criteria.

1. Children at cycle 1 must be between 2 and 5 years old.

2. The longitudinal weight at cycle 1 is not zero.

3. The survey response pattern from cycle 1 to cycle 4 is monotonic, that is, all cases have the following possible response patterns: RRRR, RRRN, RRNN, or RNNN, where R and N stand for response and nonresponse, respectively.

4. The children stay in the same province over the four cycles.

The SSC dataset consists of 1033 cases, though we have missing information for many children for at least one variable under study. Since the proposed statistical methodology described in Chapter 2 cannot handle missing data, we consider only the cases for which complete information is available for the study variables of interest (see Section 3.3 for a description of the variables under study). This leads to a reduced sample of 645 cases, which is about 62% of the total number of cases.

A comparison of the response variables (i.e., anxiety and aggression) between the full dataset and the reduced dataset is presented in Table 3.1.

Table 3.1: Comparison of the response variables between the full dataset and the reduced dataset

Responses	Full data (n=1033)			Reduced data with no missing observation (n=645)		
Anxiety	Mean	Median	SD	Mean	Median	SD
Cycle 1	0.57	0	1.087	0.87	0	1.316
Cycle 2	0.77	0	1.407	0.79	0	1.363
Cycle 3	0.96	0	1.678	1.03	0	1.689
Cycle 4	2.38	2	2.255	2.33	2	2.259
Aggression	Mean	Median	SD	Mean	Median	SD
Cycle 1	1.56	1	1.816	3.46	3	3.164
Cycle 2	1.50	1	1.782	1.63	1	1.808
Cycle 3	1.41	1	1.827	1.52	1	1.861
Cycle 4	1.28	1	1.809	1.36	1	1.796

From Table 3.1 we see that the means of the anxiety scores are similar between the full dataset and the reduced dataset for all four cycles, and the medians are exactly the same for the two data sets. The standard deviations of the anxiety scores between the two data sets are also similar. For aggression, the means, medians and the standard deviations of the scores are similar for all but cycle 1. For cycle 1, the summary statistics for aggression scores differ slightly between the two data sets. Overall, in terms of the summary statistics, the distributions of the response variables between the two data sets seem to be similar within each cycle.

We also make a comparison between the full dataset and the reduced dataset with respect to the age and gender distributions in Table 3.2. We notice a very similar gender distribution in the two datasets: about 51% males and 49% females. We also observe similar age distributions in the two datasets with the median ages being the same in each cycle and the mean ages differ very slightly. So, the dataset without missing observations may be considered representative of the complete dataset with respect to the demographic characteristics (at least for age and gender). We consider the dataset without missing observations for our analysis, and we expect our findings to be minimally affected due to ignoring the missing observations in the analysis.

Table 3.2: Comparison of demographic characteristics between the full dataset and the reduced dataset

Characteristics	Full dataset (n=1033)			Reduced data with no missing observations (n=645)		
Gender Distribution						
Male		50.7%			50.5%	
Female		49.3%			49.5%	
Age Distribution	Mean	Median	SD	Mean	Median	SD
Cycle 1	3.45	3	1.121	3.45	3	1.107
Cycle 2	5.42	5	1.113	5.42	5	1.093
Cycle 3	7.36	7	1.120	7.36	7	1.101
Cycle 4	9.39	9	1.176	9.40	9	1.163

3.3 Study Variables

Recall that our objective is to infer about the behavioral development of Canadian children over time. To this end, we consider two response variables in this study to characterize this development, namely, anxiety and aggression. Moreover, we consider four covariates to investigate their effects on the behavioral development of children, which are gender, family status, depression of person most knowledgeable and number of siblings. In the following sections, we describe these variables using frequency distributions and summary statistics.

3.3.1 Response Variables

Based on some relevant information (see below) about anxiety and aggression, factor analyses were carried out to assign scores for each of these variables for each child (Hotton, 2003; Government of Canada, 2008). The scores were assigned in such a way that a high score indicates the presence of behaviors associated with, for example, anxiety, though no cutoff value was suggested to define the high and low levels of the condition.

The anxiety scores were calculated based on the emotion levels of the children, including information about unhappiness, depression, worries, and nervousness. A score ranging from 0 to 12 was created for children of ages 2 to 3 years old, whereas the range was from 0 to 14 for children in the 4-11 age group. Since a high score indicates high level of anxiety, a reasonable choice for the cutoff value for dichotomization is to consider a high percentile value, the choice of which is somewhat arbitrary. Hotton (2003) considered 80^{th} percentile to dichotomize aggression for the whole Statistics Canada dataset. Following the recommendation by Hotton (2003), we used 80^{th} percentile to dichotomize anxiety. So, a child with a score above the 80^{th} percentile falls into the category of high level of anxiety.

The aggression score was created based on a series of questions about the frequencies of involvement of a child in aggressive activities, including fighting, bumping, attacking, bullying and threatening. A score ranging from 0 to 16 was created for children of ages 2 to 3 years old, whereas the range was from 0 to 12 for children in the 4-11 age group. The 80^{th} percentile is used to categorize aggression into high and low levels, with a score above the 80^{th} percentile is considered high level for aggression.

Table 3.3 displays the cutoff values used to define the low and high levels of anxiety and aggression. It also summarizes the distribution of the levels of these two variables. Note that we used the 80^{th} percentile as cutoff point to dichotomize the response variables. Although the choice of the 80^{th} percentile is reasonable as described above, the results presented in this thesis should be used with caution, as other cutoff values may lead to different conclusions.

Table 3.3: Distribution of the levels of anxiety and aggression

Response	Low-level	High-level	80^{th} Percentile (the cutoff value for dichotomization)
Anxiety			
Cycle 1	491 (76%)	154 (24%)	2
Cycle 2	514 (80%)	131 (20%)	2
Cycle 3	485 (75%)	160 (25%)	2
Cycle 4	489 (76%)	156 (24%)	4
Aggression			
Cycle 1	487 (76%)	158 (24%)	6
Cycle 2	482 (75%)	163 (25%)	3
Cycle 3	506 (78%)	139 (22%)	3
Cycle 4	514 (80%)	131 (20%)	3

3.3.2 Covariates

Data file used in this study has limited number of covariates that might have impacts on behavioral development of children. Moreover, the existence of a high percentage of missing values restricted us to investigate all the covariates in this study. Therefore, we settle down with four covariates which lead to a reasonable sample size of 645 children. Below, we present a description of the covariates under study along with their distribution over the four cycles.

Depression of the PMK: As part of the parent's questionnaire in NLSCY, the person most knowledgeable (PMK) about a child was asked to report his/her own depression status. A shorter version of depression rating scale of Radloff (1977) was used to measure the depression score of the PMK (Statistics Canada, 1999b). In the NLSCY each PMK was asked to reply a set of 12 questions relating to his/her loneliness, enjoyment, emotions (Statistics Canada, 1994, 2007). A depression score ranging from 0 to 36 was calculated to represent the depression scale of the PMK. A high score indicates presence of depression symptoms. We use the 80^{th} percentile of the depression score to dichotomize the depression level of PMK. We use the following coding for depression of PMK in our analysis.

$$\text{Depression of PMK} = \begin{cases} 1 & \text{if high} \\ 0 & \text{if low} \end{cases}.$$

Low level of depression of PMK is used as reference group in the data analysis.

Gender: Gender of a child is considered an important factor for anxiety and aggression (Archer and Coyne, 2005; Hotton, 2003). We investigate the effects of gender on anxiety and aggression in this study. Female children are considered as reference group. We use the following coding for gender in our analysis.

$$\text{Gender} = \begin{cases} 1 & \text{if male} \\ 0 & \text{if female} \end{cases}.$$

Number of siblings: Information was collected regarding total number of siblings (including full, half, step, adopted and foster siblings) of a child living in the same household. We use the following coding for number of siblings in our analysis.

$$\text{Number of siblings} = \begin{cases} 1 & \text{if at least one sibling} \\ 0 & \text{if no sibling} \end{cases}.$$

A child without siblings is considered as reference group to calculate the odds ratios.

Family Status: Parenting practice is considered an important factor for childhood development. We investigate the effects of child's living status on anxiety and aggression in this study. We use the following coding for family status in our analysis.

$$\text{Family Status} = \begin{cases} 1 & \text{if living with both biological parents} \\ 0 & \text{if others} \end{cases}.$$

A child not living with both biological parents is considered as reference group to calculate the odds ratios.

Table 3.4: Distribution of covariates

Variable	Cycle 1	Cycle 2	Cycle 3	Cycle 4
Depression of PMK				
Low	505 (78%)	510 (79%)	493 (76%)	507 (79%)
High	140 (22%)	135 (21%)	152 (24%)	138 (21%)
Gender				
Male	326 (51%)	326 (51%)	326 (51%)	326 (51%)
Female	319 (49%)	319 (49%)	319 (49%)	319 (49%)
Number of Siblings				
At least one sibling	532 (82%)	564 (87%)	563 (87%)	567 (88%)
No sibling	113 (18%)	81 (13%)	82 (13%)	78 (12%)
Family Status				
Both biological parents	556 (86%)	534 (83%)	517 (80%)	498 (77%)
Other	89 (14%)	111 (17%)	128 (20%)	147 (23%)

Table 3.4 presents the frequency distribution of the covariates used in our analysis. Distribution of depression level of PMK is found stable for all four cycles. Number of children not living with both biological parents is almost double in cycle 4 (147) as compared to the same in cycle 1 (89).

3.4 Analysis and Results

We summarize the fits of the marginal regression models in Table 3.5. We see that none of the covariates has significant effects on the anxiety of children in both directions (i.e., low $\rightarrow$ high and high $\rightarrow$ low) of transitions. Meanwhile, family status (p-value $= 0.075$) and depression of PMK (p-value $= 0.062$) were found statistically significant at a threshold greater than p-value < 0.05 but less than p-value < 0.10.

Table 3.5: Estimates of the regression coefficients for the marginal transition models

Transition	Covariates	Anxiety			Aggression		
		Estimate	$\hat{\text{OR}}$	95% CI	Estimate	$\hat{\text{OR}}$	95% CI
Low → High	Depression of PMK	-0.1212	0.89	(0.725, 1.082)	-0.0652	0.94	(0.731, 1.200)
	Gender	0.0430	1.04	(0.851, 1.280)	-0.3987	**0.67**	(0.512, 0.880)
	Number of siblings	-0.1904	0.83	(0.631, 1.084)	-0.0179	0.98	(0.645, 1.495)
	Family status	0.2413	1.27	(0.977, 1.658)	0.2001	1.22	(0.872, 1.712)
High → Low	Depression of PMK	0.2730	1.31	(0.988, 1.748)	0.2286	1.26	(0.995, 1.587)
	Gender	0.0413	1.04	(0.799, 1.358)	0.3700	**1.45**	(1.150, 1.823)
	Number of siblings	-0.0492	0.95	(0.582, 1.556)	0.1462	1.16	(0.841, 1.592)
	Family status	-0.0631	0.94	(0.645, 1.365)	0.0469	1.05	(0.750, 1.464)

Gender of a child has significant effects on aggressive behavior in both directions (i.e., low → high and high → low transitions). For example, the estimate of adjusted odds ratio of gender for transition from low → high level of aggression is 0.67, i.e., male children are 33% less likely to transit from low → high level of aggression compared to female children. Also, the estimate of the adjusted odds ratio for high → low transition is 1.45, which implies that a transition from high to low level of aggression is 1.45 times as likely to occur among males than females in the study population. Depression of PMK is found marginally significant with 95% CI [0.995, 1.587] for high → low transition of Aggression.

Estimates of the adjusted odds ratios along with 95% confidence intervals for the association parameters are presented in Table 3.6. We see from Table 3.6 that all four estimates of association parameters are significant at 5% level of significance, indicating that the two responses, anxiety and aggression, are correlated in this sample of children. For transition from low to high level of anxiety and aggression the estimated odds ratio is $\hat{\text{OR}} = 2.52$, i.e., children who transition from low to high level of anxiety are 2.52 times more likely to be among those children who also transit from low to high level for aggression, compared to children who do not experience transition from low level of aggression. For transition from low to high level of anxiety and from high to low level of aggression the estimated odds ratio is $\hat{\text{OR}} = 0.55$, i.e., transition out of the low level of anxiety is 0.45 times less likely to occur for those who experience a transition out of high level of aggression than who did not experience a transition out of high level of aggression. Similarly, we can describe other odds ratios.

Table 3.6: Estimates of the association parameters

Parameters	Description	$\hat{\text{OR}}$	SE	Confidence interval	
				Lower 95%	Upper 95%
$\gamma_{11}^{12} = \exp(\psi_0)$	$\frac{\text{odds}\left(N_{\text{low}}^{(\text{anxiety})} \mid N_{\text{low}}^{(\text{aggression})}=1\right)}{\text{odds}\left(N_{\text{low}}^{(\text{anxiety})} \mid N_{\text{low}}^{(\text{aggression})}=0\right)}$	2.52	0.1334	1.9443	3.2800
$\gamma_{22}^{(12)} = \exp(\psi_0 + \psi_3)$	$\frac{\text{odds}\left(N_{\text{high}}^{(\text{anxiety})} \mid N_{\text{high}}^{(\text{aggression})}=1\right)}{\text{odds}\left(N_{\text{high}}^{(\text{anxiety})} \mid N_{\text{high}}^{(\text{aggression})}=0\right)}$	2.20	0.1724	1.5666	3.0795
$\gamma_{12}^{(12)} = \exp(\psi_0 + \psi_1)$	$\frac{\text{odds}\left(N_{\text{low}}^{(\text{anxiety})} \mid N_{\text{high}}^{(\text{aggression})}=1\right)}{\text{odds}\left(N_{\text{low}}^{(\text{anxiety})} \mid N_{\text{high}}^{(\text{aggression})}=0\right)}$	0.55	0.1789	0.3875	0.7816
$\gamma_{21}^{(12)} = \exp(\psi_0 + \psi_2)$	$\frac{\text{odds}\left(N_{\text{high}}^{(\text{anxiety})} \mid N_{\text{low}}^{(\text{aggression})}=1\right)}{\text{odds}\left(N_{\text{high}}^{(\text{anxiety})} \mid N_{\text{low}}^{(\text{aggression})}=0\right)}$	0.58	0.2217	0.3777	0.9005

3.5 Discussion

In this study we consider four covariates e.g., gender, depression of person most knowledgeable, number of siblings and family status, to investigate their effects on transition behavior of anxiety and aggression among children over time. For anxiety, all four covariates are found statistically not significant for both directions of transition of low to high or high to low. Depression of PMK is found marginally significant for transition from high level of anxiety. For aggression, gender of the child is found statistically significant for both types of transition. Depression of PMK is found marginally significant for transition from high level of aggression. In addition to marginal covariates effects, the sample data suggest statistical significant relationship among all four association transitions, and indicate positive relationship between anxiety and aggression over time. In Chapter 4 we present the major finding of this thesis alongwith discussion on related studies.

Chapter 4

Discussion

At the beginning of this chapter we discuss the importance of studying two responses, anxiety and aggression, followed by discussion on findings of our analysis. We discuss few cautionary remarks about the present study in Section 4.1 and conclude this chapter with an outline to prospective future work in Section 4.2.

Early childhood experience and the environment children grow-up play a pivotal roles in future development of the children. Anxiety and aggression are considered two major obstacles for natural development of children (Craig, 1998). There are variety of causes that might be responsible for developing anxiety and aggression among children, including gender, depression level of PMK, number of siblings, child's biological parents status etc. It is of paramount importance to identify the factors that might have significant effects on anxiety and aggression, as controlling of such factors at the family and social levels would help natural growth of children.

Child's anxiety and aggression are positively correlated (Wu, Liu and Liu, 2009). In addition to psychological disorientation, anxiety is believed to have negative physical impacts on the victims (Bardone et al., 1998). As reported by Craig (1998), odds of being victimized are higher among children with anxiety disorder. Besides, victimized children naturally suffer from more health-related problems compared to non-victimized ones in school (Rigby, 2007). Moreover, childhood anxiety is highly correlated with panic disorders in adulthood (Pollack et al., 1996). According to Bittner et al. (2007), children with anxiety problems in their childhood are more likely to develop psychiatric disorders in adolescence. Previous studies also showed that children with anxiety disorders are less likely to show their proper skills (Tramonte and Willms, 2010) in different activities at school, family and society level.

Aggression generally involves the intention to cause harm on others (Archer and Coyne, 2005). Study of physical aggression in childhood is important as aggressive children are at high risk for subsequent antisocial activities (Keenan et al., 2002). Children with aggressive behavior generally show poor performance in schools and end up with unemployment, poverty and addiction (Broidy et al., 2003). Study suggests that aggression has a relation with socioeconomic status and family income (Tremblay et al., 2006).

In 2006-2007, among the Canadian children from 2 to 5 years of age, about 14.3% displayed signs associated with emotional disorder-anxiety and 12.5% suffered from being aggressive during their childhood (Government of Canada, 2011). Emotional disorders like anger, sadness, fear, grief often originate in the early years of children and are signs of development of anxiety, aggression and other emotional disorders (Ministry of Health and Long-Term Care Report, Ontario, 2011). Incidence of mental health difficulties and illnesses among children are predicted to increase by 50% by the year 2020 (Canadian Paediatric Society, 2009). Children with anxiety and aggression disorders in early ages may become problematic for a society and with early intervention, most of the emotional disorders can be prevented (Ialongo, Vaden-Kieman and Kellam, 1998). Identification of the causes behind childhood emotional disorders like anxiety and aggression plays an important role in the development of children as well as the betterment of the society.

Study suggests that female children are more aggressive compared to male (Crick and Grotpeter, 1995). According to table (3.5) female children are about 1.49 times more likely to transit from low to high level of aggression compared to male children. For anxiety, the sample dataset did not show any significant association between gender and transitions between levels of anxiety. Parenting practices are considered an important factor for behavioral development of children and a child living with a single parent is more likely to develop emotional and behavioral problems compared to child living with both biological parents (Tremblay et al., 2006). This study does not show any statistical significant association between child's family status and transition behavior for anxiety and aggression. Similarly, we did not find any significant effects of number of siblings on transitional behavior of children over time. Children of depressed parents are at high risk of developing behavioral problems compared to children of normal parents (Cummings and Davies, 1994; Weissman et al., 1997). Also, Hotton (2003) reported significant association between depression level of PMK and emotional disorder-anxiety of children. The dataset used in this study do not show any statistical significant association between depression level of PMK and transitional behavior of two responses among children. Among the four covariates considered in this study, gender of the child is only found statistically significant for low to high and high to low level of transition for aggression.

4.1 Cautionary Remarks

Some caution about the analysis and interpretation performed in this thesis is required for the following reasons.

1. Missing data are very common in longitudinal studies (Wu, Liu and Liu, 2009; Fitzmaurice, Laird and Ware, 2011). For data missing completely at random (MCAR), the GEE techniques lead to consistent estimators of population averaged prevalence (Liang and Zeger, 1986). In addition, Zeng and Cook (2007) have shown that the empirical bias for estimates of marginal and association parameters is small when observations are even missing at random (MAR). The percentage of missing information is very high in the dataset we used in this thesis, both in responses and in covariates. Assuming the missing observations as missing completely at random, we consider a reduced dataset in this study by removing all individuals with missing information in at least one of the variables under study.

2. According to the Statistics Canada (2007), a youth is considered to have behavioral problem if his/her score falls in the highest 10% of behavioral scale. Hotton (2003) considered the highest 20% scores as a definition for high level of aggression. There is no clear definition about the cutoff value to define high and low levels of anxiety and aggression. In this thesis, we consider 80^{th} percentile of the anxiety/aggression scores to define high and low levels; children with scores above the 80^{th} percentile are considered to have high anxiety/aggression level.

3. Analyses in this thesis are based on the synthetic dataset of NLSCY. Although our analysis reveals some important information including the possibility of the use of

transitional models for multivariate longitudinal binary data in the context of behavioral development of Canadian children over time, we should be cautious about drawing scientific conclusion for the population of children in Canada based on this study results.

4.2 Future Work

The proposed method of Zeng and Cook (2007) facilitates to estimate the effects of covariates on transition probabilities for multivariate longitudinal binary responses. It can help us to understand how association between two or more processes changes over time. Moreover, estimates obtained using joint transition models are more efficient compared to estimates of separate analysis of processes (Zeng and Cook, 2007). Now we outline few future work relating to this thesis.

1. Instead of binary responses, we can consider multinomial responses to investigate the effects of covariates on transitional behavior of children. We may consider three categories for each response variables, e.g., low, medium and high levels of anxiety and aggression. By considering three categories for each response variable, we expect to use more information contained in the dataset. Proposed methods of Zeng and Cook (2007) allow us to incorporate such multinomial responses alongwith higher-order Markov processes, which involves high dimensions of mathematical derivations.

2. In addition to estimating time-independent association parameters, we can consider time-dependent covariates to estimate the effects of them on association parameters. We can consider time-dependent covariates in model (2.7) and fit the regression models of conditional odds ratios, $\gamma_{i;ll'}^{(jj')}(t_k)$, for two processes j and j' as

$$\log\left(\gamma_{i;ll'}^{(jj')}\right) = \mathbf{z}_{i;ll'}^{(jj')}(t_k)'\boldsymbol{\psi}$$

   where $\mathbf{z}_{i;ll'}^{(jj')}(t_k)$ is a vector of covariates that alter the association between two process and $\boldsymbol{\psi}$ is the vector of time-dependent association parameters.

3. As mentioned in previous section that we used synthetic dataset of NLSCY for our analysis. As a result we can not draw conclusion about the children population in

Canada. Our programming codes for this thesis can be used to the original dataset of NLSCY to estimate marginal and association transitional parameters for anxiety aggression among children in Canada.

References

Albert, P. S. (2000). A transitional model for longitudinal binary data subject to nonignorable missing data. *Biometrics*, 56(2), 602–608.

Apter, A., Van Praag, H. M., Plutchik, R., Sevy, S., Korn, M., and Brown, S. (1990). Interrelationships among anxiety, aggression, impulsivity, and mood: A serotonergically linked cluster? *Psychiatry Research*, 32(2), 191–199.

Archer, J., and Coyne, S. M. (2005). An integrated review of indirect, relational, and social aggression. *Personality and Social Psychology Review*, 9(3), 212–230.

Atlas, R. S., and Pepler, D. J. (1998). Observations of bullying in the classroom. *The Journal of Educational Research*, 92(2), 86–99.

Bardone, A. M., Moffitt, T. E., Caspi, A., Dickson, N., Stanton, W. R., and Silva, P. A. (1998). Adult physical health outcomes of adolescent girls with conduct disorder, depression, and anxiety. *Journal of the American Academy of Child & Adolescent Psychiatry*, 37(6), 594–601.

Barrett, P. M., Rapee, R. M., Dadds, M. M., and Ryan, S. M. (1996). Family enhancement of cognitive style in anxious and aggressive children. *Journal of abnormal child psychology*, 24(2), 187-203.

Bittner, A., Egger, H. L., Erkanli, A., Jane Costello, E., Foley, D. L., and Angold, A. (2007). What do childhood anxiety disorders predict? *Journal of Child Psychology and Psychiatry*, 48(12), 1174–1183.

Broidy, L. M., Nagin, D. S., Tremblay, R. E., Bates, J. E., Brame, B., Dodge, K. A., Fergusson, D., Horwood, J. L., Loeber, R., and Laird, R. (2003). Developmental trajectories of childhood disruptive behaviors and adolescent delinquency: A six-site, cross-national study. *Developmental Psychology*, 39(2), 222.

Cameron, R., Brown, K. S., Best, J. A., Pelkman, C. L., Madill, C. L., Manske, S. R., and Payne, M. E. (1999). Effectiveness of a social influences smoking prevention program as a function of provider type, training method, and school risk. *American Journal of Public Health*, 89(12), 1827–1831.

Canadian Institute for Health Information (CIHI) (2009). *Youth Health Outcomes and Behaviours in Relation to Developmental Assets*. `https://secure.cihi.ca/free_products/cphi_youth_health_outcomes_aib_e.pdf`.

Canadian Paediatric Society (2009). *Are we doing enough? Status Report on Canadian Public Policy and Child and Youth Health*. `http://www.cps.ca/advocacy/StatusReport2009.pdf`.

Canadian Paediatric Society (2012). *Are we doing enough? Status Report on Canadian Public Policy and Child and Youth Health.* http://www.cps.ca/advocacy/StatusReport2012.pdf.

Carey, V., Zeger, S. L., and Diggle, P. (1993). Modelling multivariate binary data with alternating logistic regressions. *Biometrika*, 80(3), 517–526.

Carrillo, I., Chu, C., Su, W., and Xie, X. (2005). A longitudinal study of factors affecting childrens behaviour. Proceedings of the Survey Methods Section, *Statistical Society of Canada.*

Chapman, A. B., Guay-Woodford, L. M., Grantham, J. J., Torres, V. E., Bae, K. T., Baumgarten, D. A., Kenney, P. J., King, B. F., Glockner, J. F., Wetzel, L. H., Brummer, M. E., ONeill, W. C., Robbin, M. L., Bennett, W. M., Klahr, S., Hirschman, G. H., Kimmel, P. L., Thompson, P. A., and Miller, J. P. (2003). Renal structure in early autosomal-dominant polycystic kidney disease (ADPKD): the Consortium for Radiology Imaging Study of Polycystic Kidney Disease (CRISP) cohort. *Kidney International*, 64, 1035–1045.

Chorpita, B. F., and Barlow, D. H. (1998). The development of anxiety: The role of control in the early environment. *Psychological Bulletin*, 124(1), 3.

Cox, D. R. (1972). The analysis of multivariate binary data. *Applied Statistics*, 21, 113-120.

Craig, W. M. (1998). The relationship among bullying, victimization, depression, anxiety, and aggression in elementary school children. *Personality and Individual Differences*, 24(1), 123–130.

Craig, W. M., and Pepler, D. J. (1998). Observations of bullying and victimization in the school yard. *Canadian Journal of School Psychology*, 13(2), 41–59.

Craig, W. M., Pepler, D., and Atlas, R. (2000). Observations of bullying in the playground and in the classroom. *School Psychology International*, 21(1), 22–36.

Craig, W. M., and Konarski, R. (1998). *Bullying and victimization among canadian school children*, Human Resources Development Canada, Applied Research Branch. http://www.bbbst.com/site-bbbs/media/toronto/Bullying%20and%20victimizaton%20among%20Canadian%20school%20children.pdf.

Cram, S. (2004). *What is synthetic file?* DLI Update, Volume 7, Issue 1. http://www.statcan.gc.ca/dli-ild/doc/update2004-miseajour2004-eng.pdf.

Crick, N. R., and Grotpeter, J. K. (1995). Relational aggression, gender, and socialpsychological adjustment. *Child Development*, 66(3), 710–722.

Crick, N. R., Ostrov, J. M., Burr, J. E., Cullerton-Sen, C., Jansen-Yeh, E., and Ralston, P. (2006). A longitudinal study of relational and physical aggression in preschool. *Journal of Applied Developmental Psychology*, 27(3), 254–268.

Cummings, E. M., and Davies, P. T. (1994). Maternal depression and child development. *Journal of Child Psychology and Psychiatry*, 35(1), 73–122.

Dauvergne, M., & Johnson, H. (2001). *Children witnessing family violence*, Canadian Centre for Justice Statistics.

Diggle, P. (1992). Discussion of paper by Liang, K. Y., Zeger, S. L., and Qaqish, B. Multivariate regression analyses for categorical data. *Journal of the Royal Statistical Society*, B 54, 28–29.

Diggle, P., Heagerty, P., Liang, K. Y., and Zeger, S. (2002). *Analysis of longitudinal data.* 2nd ed., Oxford University Press.

Dobson, A. J. and Barnett, A. G. (2008). *An introduction to generalized linear models*, CRC press.

Dudani, A., Macpherson, A., and Tamim, H. (2010). Childhood behavior problems and unintentional injury: A longitudinal, population-based study. *Journal of Developmental & Behavioral Pediatrics*, 31(4), 276–285.

El-Sheikh, M., Cummings, E. M., and Goetsch, V. L. (1989). Coping with adults' angry behavior: Behavioral, physiological, and verbal responses in preschoolers. *Developmental Psychology*, 25(4), 490.

Fitzmaurice, G., Davidian, M., Verbeke, G., and Molenberghs, G. (Eds.)(2008). *Longitudinal Data Analysis*. Chapman and Hall/CRC Press.

Fitzmaurice, G. M., Laird, N. M., and Rotnitzky, A. G. (1993). Regression models for discrete longitudinal responses. *Statistical Science*, 284-299.

Fitzmaurice, G. M., Laird, N. M., and Ware, J. H. (2011). *Applied Longitudinal Analysis*, John Wiley & Sons.

Government of Canada (2008). *The Well–Being of Canada's Young Children*. `http://publications.gc.ca/collections/collection_2009/rhdcc-hrsdc/HS1-7-2008E.pdf`.

Government of Canada (2011). *The Well-Being of Canada's Young Children*. `http://www.dpe-agje-ecd-elcc.ca/eng/ecd/well-being/sp_1027_04_12_eng.pdf`.

Hasselman, B., (2013). *Solving System of Non Linear Equations*. R programming Package '`nleqslv`'. `http://cran.fhcrc.org/web/packages/nleqslv/nleqslv.pdf`.

Hedeker, D., and Gibbons, R. D. (2006). *Longitudinal Data Analysis*, John Wiley & Sons.

Henig, R. M. (2009). *Understanding the anxious mind*, The New York Times Magazine. `http://www.nytimes.com/2009/10/04/magazine/04anxiety-t.html?pagewanted=all`.

Hotton, T. (2003). *Childhood aggression and exposure to violence in the home*, Statistics Canada. `http://www.statcan.gc.ca/pub/85-561-m/002/4193725-eng.pdf`.

Huesmann, L. R., Moise-Titus, J., Podolski, C., and Eron, L. D. (2003). Longitudinal relations between children's exposure to TV violence and their aggressive and violent behavior in young adulthood: 1977–1992. *Developmental Psychology*, 39(2), 201.

Ialongo, N. S., Vaden-Kieman, N., and Kellam, S. (1998). Early peer rejection and aggression: Longitudinal relations with adolescent behavior. *Journal of Developmental and Physical Disabilities*, 10(2), 199–213.

Jones, O., Maillardet, R., and Robinson, A. (2009). *Introduction to scientific programming and simulation using R*, CRC Press.

Kazdin, A. E. (2000). *Encyclopedia of Psychology*, Oxford University Press.

Keenan, K., Tremblay, R., Barr, R., and Peters, R. (2002). The development and socialization of aggression during the first five years of live. *Encyclopedia on Early Childhood Development*, 1–6. http://www.child-encyclopedia.com/documents/KeenanANGxp.pdf.

Koch, G. C., Landis, J. R., Freeman, J. L., Freeman, D. H., and Lehnen, R. G. (1977). A general methodology for the analysis of repeated measurements of categorical data. *Biometrics*, 33, 133–158.

Korn, E. L., and Whittemore, A. S. (1979). Methods for analyzing panel studies of acute health effects of air pollution. *Biometrics*, 35, 795–802.

Letourneau, N. L., Tramonte, L., and Willms, J. D. (2012). Maternal depression, family functioning and children's longitudinal development. *Journal of Pediatric Nursing*, 28(3), 223-234.

Liang, K., and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1), 13-22.

Lipsitz, S. R., Laird, N. M., and Harrington, D. P. (1991). Generalized estimating equations for correlated binary data: using the odds ratio as a measure of association. *Biometrika*, 78(1), 153-160.

Maindonald, J., and Braun, W. J. (2007). *Data analysis and graphics using R: An example-based approach*, Cambridge University Press.

Marsee, M. A., Weems, C. F., and Taylor, L. K. (2008). Exploring the association between aggression and anxiety in youth: A look at aggressive subtypes, gender, and social cognition. *Journal of Child and Family Studies*, 17(1), 154-168.

McLean, C. P., Asnaani, A., Litz, B. T., and Hofmann, S. G. (2011). Gender differences in anxiety disorders: Prevalence, course of illness, comorbidity and burden of illness. *Journal of Psychiatric Research*, 45(8), 1027–1035.

Meade, M. O., Cook, R. J., Guyatt, G. H., Groll, R., Kachura, J. R., Bedard, M., Cook, R. J., Slutsky, A. S., and Stewart, T. E. (2000). Interobserver variation in interpreting chest radiographs for the diagnosis of acute respiratory distress syndrome. *American Journal of Respiratory and Critical Care Medicine*, 161(1), 85–90.

Ministry of Health and Long-Term Care Report (2011), Ontario. *Open Minds, Healthy Minds. Ontario's Comprehensive Mental Health and Addictions Strategy.* http://www.health.gov.on.ca/en/common/ministry/publications/reports/mental_health2011/mentalhealth_rep2011.pdf.

Moss, K. (2003). *Witnessing violence-aggression and anxiety in young children.* Health reports, Statistics Canada, Canadian Centre for Health Information, 14, 53.

Neuhaus, J. M. (1992). Statistical methods for longitudinal and clustered designs with binary responses. *Statistical Methods in Medical Research*, 1(3), 249–273.

Ochi, Y., and Prentice, R. L. (1984). Likelihood inference in a correlated probit regression model. *Biometrika*, 71(3), 531–543.

Parens, H. (2008). *The development of aggression in early childhood*, Revised ed., Jason Aronson.

Pepler, D. J., and Sedighdeilami, F. (1998). *Aggressive girls in Canada*, Human Resources Development Canada, Applied Research Branch.

Pollack, M. H., Otto, M. W., Sabatino, S., Majcher, D., Worthington, J. J., McArdle, E. T., and Rosenbaum, J. F. (1996). Relationship of childhood anxiety to adult panic disorder: Correlates and influence on course. *The American Journal of Psychiatry*, 153(3), 376–381.

Prochaska, J. O., and DiClemente, C. C. (1983). Stages and processes of self-change of smoking: Toward an integrative model of change. *Journal of Consulting and Clinical Psychology*, 51(3), 390.

R Core Team (2013). R: A language and environment for statistical computing. `http://www.R-project.org`.

Radloff, L. S. (1977). The CES-D scale a self-report depression scale for research in the general population. *Applied psychological measurement*, 1(3), 385–401.

Rigby, K. (2007). *Bullying in schools and what to do about it: Revised and updated*. Australian Council for Education Research.

Ross, D. P., and Roberts, P. (1999). *Income and child well-being: A new perspective on the poverty debate*, Canadian Council on Social Development, Ottawa.

Statistics Canada (1994). *National Longitudinal Survey of Children and Youth, Cycle 1-microdata user guide*. `http://www23.statcan.gc.ca/imdb-bmdi/document/4450_D4_T9_V1-eng.pdf`.

Statistics Canada (1999a). *National Longitudinal Survey of Children and Youth: transition into adolescence, 1996/97*. July 6, 1999. `http://www.statcan.gc.ca/daily-quotidien/990706/dq990706a-eng.htm`.

Statistics Canada (1999b). *National Longitudinal Survey of Children and Youth Cycle 2 Release 3-Primary File*, December 17, 1999. `http://www.statcan.gc.ca/rdc-cdr/pdf/nlscy1996-elnej1996-5-eng.pdf`.

Statistics Canada (2007). *National Longitudinal Survey of Children and Youth, Cycle 7-microdata user guide*. `http://www23.statcan.gc.ca/imdb-bmdi/document/4450_D4_T9_V7-eng.pdf`.

Statistics Canada (2008). *National Longitudinal Survey of Children and Youth (2008)*, `http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=4450`.

Statistical Society of Canada (2005). `http://www.ssc.ca/archive/documents/case_studies/2005/nlscy_student_e.html`.

Stiratelli, R., Laird, N., and Ware, J. H. (1984). Random-effects models for serial observations with binary response. *Biometrics*, 40, 961-971.

Tattum, D. P., and Lane, D. A. (Eds.) (1989). *Bullying in schools*, Trentham Books.

Thiebaut R, Jacqmin-Gadda H, Chene G, Leport C and Commenges D (2002). Bivariate linear mixed models using SAS proc MIXED. *Computer Methods and Programs in Biomedicine*, 69: 249–256.

Tramonte, L., and Willms, D. (2010). The prevalence of anxiety among middle and secondary school students in Canada. *Canadian Journal of Public Health*, 101(9), 19-22.

Tremblay, R. E. (2002). Prevention of injury by early socialization of aggressive behavior. *Injury Prevention*, 8(suppl 4), iv17–iv21.

Tremblay, R. E., Côté, S., Vaillancourt, T., LeBlanc, J. C., and Nagin, D. S. (2006). The development of physical aggression from toddlerhood to pre-adolescence: A nation wide longitudinal study of canadian children. *Journal of Abnormal Child Psychology*, 34(1), 68–82.

Waddell, C., McEwan, K., Shepherd, C. A., Offord, D. R. and Hua, J. M. (2005). A public health strategy to improve the mental health of canadian children. *Canadian Journal of Psychiatry*, 50, 226-233.

Wedderburn, R. W. M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika*, 61, 439-447.

Wagner, M. M. (1995). Outcomes for youths with serious emotional disturbance in secondary school and early adulthood. *The Future of Children*, 5(2), 90–112.

Ware, J. H., Lipsitz, S. R., and Speizer, F. E. (1988). Issues in the analysis of repeated categorical outcomes. *Statistics in Medicine*, 7, 95-107.

Weissman, M. M., Warner, V., Wickramaratne, P., Moreau, D., and Olfson, M. (1997). Offspring of depressed parents: 10 years later. *Archives of General Psychiatry*, 54(10), 932–940.

Williams, P. (2011). *Aggression: From fantasy to action*, Karnac Books.

Wu, L., Liu, W., and Liu, J. (2009). A longitudinal study of children's aggressive behaviours based on multivariate mixed models with incomplete data. *Canadian Journal of Statistics*, 37(3), 435–452.

Zalta, A. K., and Chambless, D. L. (2012). Understanding gender differences in anxiety the mediating effects of instrumentality and mastery. *Psychology of Women Quarterly*, 36(4), 488–499.

Zeger, S. L., Liang, K. Y., and Albert, P. S. (1988). Models for longitudinal data: A generalized estimating equation approach. *Biometrics*, 44(4), 1049–1060.

Zeger, S. L., Liang, K. Y., and Self, S. G. (1985). The analysis of binary longitudinal data with time-independent covariates. *Biometrika*, 72, 31-38.

Zeng, L., and Cook, R. J. (2007). Transition models for multivariate longitudinal binary data. *Journal of the American Statistical Association*, 102(477), 211–223.

Appendix A

Derivation of Mathematical Expressions for Marginal Models

In this chapter, we present the detailed derivation of mathematical expressions related to marginal models.

A.1 Odds Ratios, $\gamma_{i;ll'}^{(jj')}(t_k)$

Here we show the derivation of equation (2.9) of odds ratios. Using the marginal transition probabilities for both processes j and j' at time (t_k), we can define the distribution of joint transition probabilities as follows:

Table A.1: Distribution of joint transition probabilities

	Level	Process j': l	Process j': l'	Marginal
Process j	l	$\pi_{i;ll'}^{(jj')}(t_k)$	$\left(\pi_{il}^{(j)}(t_k) - \pi_{i;ll'}^{(jj')}(t_k)\right)$	$\pi_{il}^{(j)}(t_k)$
	l'	$\left(\pi_{il}^{(j')}(t_k) - \pi_{i;ll'}^{(jj')}(t_k)\right)$	$\left(1 - \pi_{il}^{(j)}(t_k) - \pi_{il'}^{(j')}(t_k) + \pi_{i;ll'}^{(jj')}(t_k)\right)$	$\pi_{il'}^{(j)}(t_k)$
Marginal		$\pi_{il}^{(j')}(t_k)$	$\pi_{il'}^{(j')}(t_k)$	1

From Table (A.1) the odds ratios can be defined using 2×2 contingency table formula of $\dfrac{ad}{bc}$ as

$$\gamma_{i;ll'}^{(jj')}(t_k) = \frac{\pi_{i;ll'}^{(jj')}(t_k)\left(1 - \pi_{il}^{(j)}(t_k) - \pi_{il'}^{(j')}(t_k) + \pi_{i;ll'}^{(jj')}(t_k)\right)}{\left(\pi_{il}^{(j)}(t_k) - \pi_{i;ll'}^{(jj')}(t_k)\right)\left(\pi_{il'}^{(j')}(t_k) - \pi_{i;ll'}^{(jj')}(t_k)\right)} \tag{A.1}$$

where

$$\begin{aligned}
a &= \pi_{i;ll'}^{(jj')}(t_k) \\
b &= \pi_{il}^{(j)}(t_k) - \pi_{i;ll'}^{(jj')}(t_k) \\
c &= \pi_{il}^{(j')}(t_k) - \pi_{i;ll'}^{(jj')}(t_k), \text{ and} \\
d &= 1 - \pi_{il}^{(j)}(t_k) - \pi_{il'}^{(j')}(t_k) + \pi_{i;ll'}^{(jj')}(t_k).
\end{aligned}$$

A.2 Conditional Joint Probabilities, $\pi_{i;ll'}^{(jj')}(t_k)$

Derivation of equation (2.10) is presented here. From (A.1), the odds ratio for conditional joint probability for pair of events $\left(N_{il}^{(j)}(t_k), N_{il'}^{(j')}(t_k)\right)$ is given as

$$\gamma_{i;ll'}^{(jj')}(t_k) = \frac{\pi_{i;ll'}^{(jj')}(t_k)\left(1-\pi_{il}^{(j)}(t_k)-\pi_{il'}^{(j')}(t_k)+\pi_{i;ll'}^{(jj')}(t_k)\right)}{\left(\pi_{il}^{(j)}(t_k)-\pi_{i;ll'}^{(jj')}(t_k)\right)\left(\pi_{il'}^{(j')}(t_k)-\pi_{i;ll'}^{(jj')}(t_k)\right)}. \tag{A.2}$$

We can write the above equation as

$$\begin{aligned}\pi_{i;ll'}^{(jj')}(t_k) - \pi_{i;ll'}^{(jj')}(t_k) \times \pi_{il}^{(j)}(t_k) - \pi_{i;ll'}^{(jj')}(t_k) \times \pi_{il'}^{(j')}(t_k) + \left(\pi_{i;ll'}^{(jj')}(t_k)\right)^2 = \gamma_{i;ll'}^{(jj')}(t_k)\times \\ \left(\pi_{il}^{(j)}(t_k)\pi_{il'}^{(j')}(t_k) - \pi_{il}^{(j)}(t_k)\pi_{i;ll'}^{(jj')}(t_k) - \pi_{i;ll'}^{(jj')}(t_k)\pi_{il'}^{(j')}(t_k) + \left(\pi_{i;ll'}^{(jj')}(t_k)\right)^2\right)\end{aligned} \tag{A.3}$$

and after simplification, we can re-write (A.3) as

$$\begin{aligned}\left\{1-\gamma_{i;ll'}^{(jj')}(t_k)\right\}\left\{\pi_{i;ll'}^{(jj')}(t_k)\right\}^2 + \left\{1-(1-\gamma_{i;ll'}^{(jj')}(t_k))(\pi_{il}^{(j)}(t_k)+\pi_{il'}^{(j')}(t_k))\right\}\pi_{i;ll'}^{(jj')}(t_k) - \\ \gamma_{i;ll'}^{(jj')}(t_k)\pi_{il}^{(j)}(t_k)\pi_{il'}^{(j')}(t_k) = 0\cdot\end{aligned} \tag{A.4}$$

Using the formula for solution of quadratic equation $ax^2+bx+c=0$, i.e., $x = \dfrac{-b\pm\sqrt{b^2-4ac}}{2a}$, we can derive $\pi_{i;ll'}^{(jj')}(t_k)$ from (A.4) as

$$\pi_{i;ll'}^{(jj')}(t_k) = \begin{cases}\left(a_{i;ll'}^{(jj')}(t_k) - \left(a_{i;ll'}^{(jj')}(t_k)^2 - 4\gamma_{i;ll'}^{(jj')}(t_k)(\gamma_{i;ll'}^{(jj')}(t_k)-1)\times\pi_{il}^{(j)}(t_k)\pi_{il'}^{(j')}(t_k)\right)^{\frac{1}{2}}\right)\times \\ \quad \left(2(\gamma_{i;ll'}^{(jj')}(t_k)-1)\right)^{-1}, \text{ if } \gamma_{i;ll'}^{(jj')}(t_k)\neq 1 \\ \pi_{il}^{(j)}(t_k)\pi_{il'}^{(j')}(t_k), \quad \text{otherwise}\end{cases}$$

where $a_{i;ll'}^{(jj')}(t_k) = 1 - \left(1-\gamma_{i;ll'}^{(jj')}(t_k)\right)\left(\pi_{il}^{(j)}(t_k)+\pi_{il'}^{(j')}(t_k)\right)$.

A.3 Conditional Expectations, $\mu_i^{(j)}(t_k)$

We can derive the equation (2.17) using the conditional expectation of $N_i^{(j)}(t_k)$ for given $\boldsymbol{\delta}_i(t_k)$ and $\mathbf{x}_i(t_k)$ as follows:

$$
\begin{aligned}
\mu_i^{(j)}(t_k) &= \mathrm{E}\left(N_i^{(j)}(t_k)|\boldsymbol{\delta}_i(t_k),\mathbf{x}_i(t_k)\right)\\
&= \mathrm{E}\left(\sum_{l=1,2}\delta_{il}^{(j)}(t_k)N_{il}^{(j)}(t_k)|\boldsymbol{\delta}_i(t_k),\mathbf{x}_i(t_k)\right)\\
&= \sum_{l=1,2}\delta_{il}^{(j)}(t_k)\ \mathrm{E}\left(N_{il}^{(j)}(t_k)\right)\\
&= \delta_{i1}^{(j)}(t_k)\mathrm{E}\left(N_{i1}^{(j)}(t_k)\right)+\delta_{i2}^{(j)}(t_k)\mathrm{E}\left(N_{i2}^{(j)}(t_k)\right)\\
&= \delta_{i1}^{(j)}(t_k)\left(N_{i1}^{(j)}(t_k)\ \mathrm{P}\left(N_{il}^{(j)}(t_k)=1\right)\right)+\delta_{i2}^{(j)}(t_k)\left(N_{i2}^{(j)}(t_k)\ \mathrm{P}\left(N_{i2}^{(j)}(t_k)=1\right)\right)\\
&= \delta_{i1}^{(j)}(t_k)\left(N_{i1}^{(j)}(t_k)\ \pi_{i1}^{(j)}(t_k)\right)+\delta_{i2}^{(j)}(t_k)\left(N_{i2}^{(j)}(t_k)\ \pi_{i2}^{(j)}(t_k)\right)\\
&= \sum_{l=1,2}\delta_{il}^{(j)}(t_k)\pi_{il}^{(j)}(t_k)\cdot
\end{aligned}
$$

A.4 Partial Derivatives of Mean Veactor, $\dfrac{\partial\boldsymbol{\mu}_i(t_k)}{\partial\boldsymbol{\beta}'}$

Here we present the derivations of partial derivatives of $\boldsymbol{\mu}_i(t_k)$ with respect to $\boldsymbol{\beta}$, i.e., $D_i(t_k;\boldsymbol{\beta})=\dfrac{\partial\boldsymbol{\mu}_i(t_k)}{\partial\boldsymbol{\beta}'}$ in equation (2.18). Let $\boldsymbol{\mu}_i(t_k)$=($\mu_i^{(1)}(t_k)$, $\mu_i^{(2)}(t_k),\cdots,\mu_i^{(J)}(t_k))'$ be the vector of mean responses. The vector of regression coefficients for all processes is defined as $\boldsymbol{\beta}=(\boldsymbol{\beta}^{(1)\prime},\boldsymbol{\beta}^{(2)\prime},\cdots,\boldsymbol{\beta}^{(J)\prime})'$. The partial derivative of $\boldsymbol{\mu}_i(t_k)$ with respect to $\boldsymbol{\beta}$, $\dfrac{\partial\boldsymbol{\mu}_i(t_k)}{\partial\boldsymbol{\beta}'}$, is a $J\times(2\sum_{l=1}^{2}p_j)$ matrix of derivatives and p_j is the number of covariates associated with the transition probabilities for process j. Partial derivatives are given as

$$
\begin{aligned}
\frac{\partial\boldsymbol{\mu}_i(t_k)}{\partial\boldsymbol{\beta}'} &= \frac{\partial\left(\mu_i^{(1)}(t_k),\ \mu_i^{(2)}(t_k),...,\ \mu_i^{(J)}(t_k)\right)'}{\partial\left(\boldsymbol{\beta}^{(1)\prime},\ \boldsymbol{\beta}^{(2)\prime},...,\ \boldsymbol{\beta}^{(J)\prime}\right)'}\\
&= \begin{pmatrix}
\frac{\partial\mu_i^{(1)}(t_k)}{\partial\boldsymbol{\beta}^{(1)\prime}}, & \frac{\partial\mu_i^{(1)}(t_k)}{\partial\boldsymbol{\beta}^{(2)\prime}}, & . \ . \ . & \frac{\partial\mu_i^{(1)}(t_k)}{\partial\boldsymbol{\beta}^{(J)\prime}}\\
\frac{\partial\mu_i^{(2)}(t_k)}{\partial\boldsymbol{\beta}^{(1)\prime}}, & \frac{\partial\mu_i^{(2)}(t_k)}{\partial\boldsymbol{\beta}^{(2)\prime}}, & . \ . \ . & \frac{\partial\mu_i^{(2)}(t_k)}{\partial\boldsymbol{\beta}^{(J)\prime}}\\
... & ... & & ...\\
\frac{\partial\mu_i^{(J)}(t_k)}{\partial\boldsymbol{\beta}^{(1)\prime}}, & \frac{\partial\mu_i^{(J)}(t_k)}{\partial\boldsymbol{\beta}^{(2)\prime}}, & . \ . \ . & \frac{\partial\mu_i^{(J)}(t_k)}{\partial\boldsymbol{\beta}^{(J)\prime}}
\end{pmatrix}.
\end{aligned}
$$

For subject i, the conditional expectation of $N_i^{(j)}(t_k)$ for process j at time t_k for given $\boldsymbol{\delta}_i(t_k)$ and $\mathbf{x}_i(t_k)$ is defined as

$$\mu_i^{(j)}(t_k) = E\left(N_i^{(j)}(t_k)|\boldsymbol{\delta}_i(t_k),\ \mathbf{x}_i(t_k)\right) = \sum_{l=1,2} \delta_{il}^{(j)}(t_k)\pi_{il}^{(j)}(t_k)$$

and the partial derivatives of (A.6) with respect to marginal transition parameters $\boldsymbol{\beta}_l^{(j)}(t_k)$ can be defined as

$$\begin{aligned}
\frac{\partial\left(\mu_i^{(j)}(t_k)\right)}{\partial\boldsymbol{\beta}_l^{(j)\prime}(t_k)} &= \frac{\partial}{\partial\boldsymbol{\beta}_l^{(j)\prime}(t_k)}\left(\sum_{l=1,2}\delta_{il}^{(j)}(t_k)\pi_{il}^{(j)}(t_k)\right)\\
&= \sum_{l=1,2}\delta_{il}^{(j)}(t_k)\frac{\partial}{\partial\boldsymbol{\beta}_l^{(j)\prime}(t_k)}\left(\pi_{il}^{(j)}(t_k)\right)\\
&= \sum_{l=1,2}\delta_{il}^{(j)}(t_k)\frac{\partial}{\partial\boldsymbol{\beta}_l^{(j)\prime}}(t_k)\left(\frac{e^{\left(\mathbf{x}_i^{(j)\prime}(t_k)\boldsymbol{\beta}_l^{(j)}(t_k)\right)}}{1+e^{\left(\mathbf{x}_i^{(j)\prime}(t_k)\boldsymbol{\beta}_l^{(j)}(t_k)\right)}}\right)\\
&= \sum_{l=1,2}\delta_{il}^{(j)}(t_k)\frac{\mathbf{x}_i^{(j)\prime}(t_k)e^{\left(\mathbf{x}_i^{(j)\prime}(t_k)\boldsymbol{\beta}_l^{(j)}(t_k)\right)}}{\left(1+e^{\left(\mathbf{x}_i^{(j)\prime}(t_k)\boldsymbol{\beta}_l^{(j)}(t_k)\right)}\right)^2}\\
&= \sum_{l=1,2}\delta_{il}^{(j)}(t_k)\left(\frac{e^{\left(\mathbf{x}_i^{(j)\prime}(t_k)\boldsymbol{\beta}_l^{(j)}(t_k)\right)}}{1+e^{\left(\mathbf{x}_i^{(j)\prime}(t_k)\boldsymbol{\beta}_l^{(j)}(t_k)\right)}}\right)\left(1-\frac{e^{\left(\mathbf{x}_i^{(j)\prime}(t_k)\boldsymbol{\beta}_l^{(j)}(t_k)\right)}}{1+e^{\left(\mathbf{x}_i^{(j)\prime}(t_k)\boldsymbol{\beta}_l^{(j)}(t_k)\right)}}\right)\mathbf{x}_i^{(j)\prime}(t_k)\\
&= \sum_{l=1,2}\delta_{il}^{(j)}(t_k)\pi_{il}^{(j)}(t_k)(1-\pi_{il}^{(j)}(t_k))\mathbf{x}_i^{(j)\prime}(t_k)\cdot \qquad \text{(A.6)}
\end{aligned}$$

A.5 Working Covariance Matrix, $v_i^{(jj')}(t_k)$

The derivation of equation (2.20) is given below. The covariance matrix, $v_i^{(jj')}(t_k)$, for process j and j' for given $\boldsymbol{\delta}_i(t_k)$, and$\mathbf{x}_i(t_k)$ is written as

$$v_i^{(jj')}(t_k) = \operatorname{cov}\left(N_i^{(j)}(t_k),\ N_i^{(j')}(t_k)|\boldsymbol{\delta}_i(t_k),\ \mathbf{x}_i(t_k)\right)\cdot$$

Thus, the covariance of $N_i^{(j)}(t_k)$ and $N_i^{(j')}(t_k)$ is given by:

$$\begin{aligned}
&\operatorname{Cov}\left(N_i^{(j)}(t_k), N_i^{(j')}(t_k)|\boldsymbol{\delta}_i(t_k), \mathbf{x}_i(t_k)\right)\\
&= \mathrm{E}\left(\left(N_i^{(j)}(t_k)-\mathrm{E}(N_i^{(j)}(t_k)\right)\ \left(N_i^{(j')}(t_k)-\mathrm{E}(N_i^{(j')}(t_k)\right)\right)\\
&= \mathrm{E}\left((N_i^{(j)}(t_k)-\mu_i^{(j)}(t_k))\ (N_i^{(j')}(t_k)-\mu_i^{(j')}(t_k))\right)\\
&= \mathrm{E}\left(N_i^{(j)}(t_k)N_i^{(j')}(t_k)-N_i^{(j)}(t_k)\mu_i^{(j')}(t_k)-\mu_i^{(j)}(t_k)N_i^{(j')}(t_k)-\mu_i^{(j)}(t_k)\mu_i^{(j')}(t_k)\right)
\end{aligned}$$

$$= \mathrm{E}\left(N_i^{(j)}(t_k)N_i^{(j')}(t_k)\right) - \mu_i^{(j')}(t_k)\mathrm{E}\left(N_i^{(j)}(t_k)\right) - \mu_i^{(j)}(t_k)\mathrm{E}\left(N_i^{(j')}(t_k)\right) + \mu_i^{(j)}(t_k)\mu_i^{(j')}(t_k)\cdot$$

Now,

$$\begin{aligned}
&\mathrm{E}\Big(N_i^{(j)}(t_k)N_i^{(j')}(t_k)|\boldsymbol{\delta}_i(t_k), \mathbf{x}_i(t_k)\Big) \\
&= \mathrm{E}\left(\sum_{l=1}^{2}\delta_{il}^{(j)}(t_k)N_{il}^{(j)}(t_k)\sum_{l'=1}^{2}\delta_{il'}^{(j')}(t_k)N_{il'}^{(j')}(t_k)\right) \\
&= \mathrm{E}\left(\sum_{l,l'=1}^{2}\delta_{il}^{(j)}(t_k)\delta_{il'}^{(j')}(t_k)N_{il}^{(j)}(t_k)N_{il'}^{(j')}(t_k)\right) \\
&= \sum_{l,l'=1}^{2}\delta_{il}^{(j)}(t_k)\delta_{il'}^{(j')}(t_k)\mathrm{E}\left(N_{il}^{(j)}(t_k)N_{il'}^{(j')}(t_k)\right) \\
&= \sum_{l,l'=1}^{2}\delta_{il}^{(j)}(t_k)\delta_{il'}^{(j')}(t_k)\left(N_{il}^{(j)}(t_k)N_{il'}^{(j')}(t_k)\mathrm{P}\left(N_{il}^{(j)}(t_k)=1,\ N_{il'}^{(j')}(t_k)=1\right)\right) \\
&= \sum_{l,l'=1}^{2}\delta_{il}^{(j)}(t_k)\delta_{il'}^{(j')}(t_k)\pi_{i;ll'}^{(jj')}(t_k)
\end{aligned}$$

where, $\pi_{i;ll'}^{(jj')}(t_k) = P\left(N_{il}^{(j)}(t_k)=1,\ N_{il'}^{(j')}(t_k)=1|\delta_{il}^{(j)}=1,\ \delta_{il'}^{(j')}=1,\ \mathbf{x}_i(t_k)\right)$ denotes the conditional joint probability for the pair of events $(N_{il}^{(j)}(t_k),\ N_{il'}^{(j')}(t_k))$.

Then,

$$\begin{aligned}
&\mathrm{cov}\Big(N_i^{(j)}(t_k), N_i^{(j')}(t_k)|\boldsymbol{\delta}_i(t_k), \mathbf{x}_i(t_k)\Big) \\
&= \sum_{l,l'=1}^{2}\delta_{il}^{(j)}(t_k)\delta_{il'}^{(j')}(t_k)\pi_{i;ll'}^{(jj')}(t_k) - \sum_{l'=1}^{2}\delta_{il}^{(j')}(t_k)\pi_{il}^{(j')}(t_k)\times\sum_{l=1}^{2}\delta_{il}^{(j)}(t_k)\pi_{il}^{(j)}(t_k) - \sum_{l=1}^{2}\delta_{il}^{(j)}(t_k)\pi_{il}^{(j)}(t_k)\times \\
&\sum_{l'=1}^{2}\delta_{il}^{(j')}(t_k)\pi_{il}^{(j')}(t_k) + \sum_{l=1}^{2}\delta_{il}^{(j)}(t_k)\pi_{il}^{(j)}(t_k)\times\sum_{l'=1}^{2}\delta_{il}^{(j')}(t_k)\pi_{il}^{(j')}(t_k) \\
&= \sum_{l,l'=1}^{2}\delta_{il}^{(j)}(t_k)\delta_{il'}^{(j')}(t_k)\pi_{i;ll'}^{(jj')}(t_k) - \sum_{l'=1}^{2}\delta_{il}^{(j')}(t_k)\pi_{il}^{(j')}(t_k)\times\sum_{l=1}^{2}\delta_{il}^{(j)}(t_k)\pi_{il}^{(j)}(t_k) \\
&= \sum_{l,l'=1}^{2}\delta_{il}^{(j)}(t_k)\delta_{il'}^{(j')}(t_k)\pi_{i;ll'}^{(jj')}(t_k) - \sum_{l,l'=1}^{2}\delta_{il}^{(j)}(t_k)\delta_{il}^{(j')}(t_k)\pi_{il}^{(j)}(t_k)\pi_{il}^{(j')}(t_k) \\
&= \sum_{l,l'=1}^{2}\delta_{il}^{(j)}(t_k)\delta_{il'}^{(j')}(t_k)\left(\pi_{i;ll'}^{(jj')}(t_k) - \pi_{il}^{(j)}(t_k)\pi_{il}^{(j')}(t_k)\right)\cdot
\end{aligned}$$

Thus the working covariance matrix of $\mathbf{N}_i(t_k)$ conditional on $\boldsymbol{\delta}_i(t_k)$ and $\mathbf{x}_i(t_k)$ can be expressed as, for $j \neq j'$

$$\begin{aligned}
&\mathrm{cov}\Big(N_i^{(j)}(t_k), N_i^{(j')}(t_k)|\boldsymbol{\delta}_i(t_k), \mathbf{x}_i(t_k)\Big) \\
&= \sum_{l,l'=1}^{2}\delta_{il}^{(j)}(t_k)\delta_{il'}^{(j')}(t_k)\left(\pi_{i;ll'}^{(jj')}(t_k) - \pi_{il}^{(j)}(t_k)\pi_{il}^{(j')}(t_k)\right)
\end{aligned}$$

and for $j = j'$,

$$v_i^{(jj')}(t_k) = \sum_{l=1,2} \delta_{il}^{(j)}(t_k) \left(\pi_{il}^{(j)}(t_k) - \left(\pi_{il}^{(j)}(t_k) \right)^2 \right)$$
$$= \sum_{l=1,2} \delta_{il}^{(j)}(t_k) \pi_{il}^{(j)}(t_k) \left(1 - (\pi_{il}^{(j)}(t_k) \right) \cdot$$

Then, the elements of covariance matrix $v_i^{jj'}$ is given as:

$$v_i^{(jj')}(t_k) = \begin{cases} \sum\limits_{l=1,2} \delta_{il}^{(j)}(t_k) \pi_{il}^{(j)}(t_k) \left(1 - (\pi_{il}^{(j)}(t_k) \right), \text{ if } j = j' \\ \sum\limits_{l,l'=1}^{2} \delta_{il}^{(j)}(t_k) \delta_{il'}^{(j')}(t_k) \left(\pi_{i;ll'}^{(jj')}(t_k) - \pi_{il}^{(j)}(t_k) \pi_{il}^{(j')}(t_k) \right), \text{ if } j \neq j' \end{cases} \cdot \qquad \text{(A.7)}$$

Appendix B

Derivation of Mathematical Expressions for Association Models

In this chapter we show derivations of mathematical expressions used for estimating association parameters.

B.1 Conditional Expectations, $\xi_i^{(jj')}(t_k)$

Here we show the derivation of equation (2.22) of $N_i^{(j)}(t_k)$ for process j given $N_i^{(j')}(t_k)$ for process j'.

$$
\begin{aligned}
\xi_i^{(jj')}(t_k) &= \mathrm{E}\left(N_i^{(j)}(t_k)|N_i^{(j')}(t_k), \boldsymbol{\delta}_i(t_k), \mathbf{x}_i(t_k)\right) \\
&= \mathrm{E}\left(\sum_{l=1}^{2} \delta_{il}^{(j)}(t_k) N_{il}^{(j)}(t_k) | \sum_{l'=1}^{2} \delta_{il'}^{(j')}(t_k) N_{il'}^{(j')}(t_k)\right) \\
&= \sum_{l,l'=1}^{2} \delta_{il}^{(j)}(t_k)\delta_{il'}^{(j')}(t_k) \mathrm{E}\left(N_{il}^{(j)}(t_k)|N_{il'}^{(j')}(t_k)\right) \\
&= \sum_{l,l'=1}^{2} \delta_{il}^{(j)}(t_k)\delta_{il'}^{(j')}(t_k) \left(N_{il}^{(j)}(t_k) \mathrm{P}\left(N_{il}^{(j)}(t_k) = 1|N_{il'}^{(j')}(t_k)\right)\right) \\
&= \sum_{l,l'=1}^{2} \delta_{il}^{(j)}(t_k)\delta_{il'}^{(j')}(t_k)\ \mathrm{P}\left(N_{il}^{(j)}(t_k) = 1|N_{il'}^{(j')}(t_k)\right) \\
&= \sum_{l,l'=1}^{2} \delta_{il}^{(j)}(t_k)\delta_{il'}^{(j')}(t_k) \xi_{i;ll'}^{(jj')}(t_k) \qquad \text{(B.1)}
\end{aligned}
$$

where $\xi_{i,ll'}^{(jj')}(t_k) = \Pr\left(N_{il}^{(j)}(t_k) = 1|N_{il'}^{(j')}(t_k), \delta_{il}^{(j)}(t_k) = 1,\ \delta_{il'}^{(j')}(t_k) = 1,\ \mathbf{x}_i(t_k)\right),\ j < j'$.

B.2 The logit Function of $\xi_i^{(jj')}(t_k)$

Let Y_{ij} and Y_{ik} are vectors of two binary responses and let γ_{ijk} be the log odds ratio between the outcomes of two responses. Let, $\mu_{ij} = \Pr(Y_{ij} = 1)$ and $\nu_{ijk} = \Pr(Y_{ij} = 1, Y_{ik} = 1)$. Then following (Carey, Zeger and Diggle, 1993; Diggle, 1992), the logit function of $\Pr(Y_{ij} = 1|Y_{ik} = y_{ik})$ can be expressed as:

$$
\operatorname{logit}\left(\Pr\left(Y_{ij} = 1|Y_{ik} = y_{ik}\right)\right) = \gamma_{ijk} y_{ik} + \log\left(\frac{\mu_{ij} - \nu_{ijk}}{1 - \mu_{ij} - \mu_{ik} + \nu_{ijk}}\right) \qquad \text{(B.2)}
$$

From (B.2), pairwise log odds ratio γ_{ijk} can be viewed as regression coefficient in a logistic regression of Y_{ij} and Y_{ik} and the second term in right-hand can be considered as offset.

From (B.2), conditional expectations, $\xi_{i;ll'}^{(jj')}(t_k)$, for pair of events $N_i^{(j)}(t_k)$ and $N_i^{(j')}(t_k)$ can be written as

$$\operatorname{logit}\left(\xi_{i,ll'}^{(jj')}(t_k)\right) = \log\left(\gamma_{i,ll'}^{(jj')}(t_k)\right) N_i^{(j')}(t_k) + \log\left(\frac{\pi_{il}^{(j)}(t_k) - \pi_{i;ll'}^{(jj')}(t_k)}{1 - \pi_{il}^{(j)}(t_k) - \pi_{il'}^{(j')}(t_k) + \pi_{i;ll'}^{(jj')}(t_k)}\right). \tag{B.3}$$

B.3 Estimation of Association Parameters, $\boldsymbol{\psi}$

Here we present the derivations required for calculating estimation equations in (2.25). The association parameter $\boldsymbol{\psi}$ can be obtained using the following estimating equations

$$\mathbf{U}_2(\boldsymbol{\beta}, \boldsymbol{\psi}) = \sum_{i=1}^{n} \sum_{k=1}^{K-1} C_i(t_k|\boldsymbol{\beta}, \boldsymbol{\psi})' S_i(t_k|\boldsymbol{\beta}, \boldsymbol{\psi})^{-1} \boldsymbol{\epsilon}_i(t_k) = 0 \tag{B.4}$$

where $C_i(t_k|\boldsymbol{\beta}, \boldsymbol{\psi}) = \dfrac{\partial \boldsymbol{\xi}_i(t_k)}{\partial \boldsymbol{\psi}'}$ is the matrix of partial derivatives of $\boldsymbol{\xi}_i(t_k) = \left(\xi_i^{(jj')}(t_k),\ j < j'\right)'$ with respect to $\boldsymbol{\psi}$ and the working independence covariance matrix of $\boldsymbol{\epsilon}_i(t_k)$ is $S_i(t_k|\boldsymbol{\beta}, \boldsymbol{\psi}) = \operatorname{diag}\left\{\xi_i^{(jj')}(t_k)\left(1 - \xi_i^{(jj')}(t_k)\right),\ j < j'\right\}$

$$\begin{aligned}
\frac{\partial \boldsymbol{\xi}_i(t_k)}{\partial \boldsymbol{\psi}'} &= \frac{\partial \xi_i^{(jj')}(t_k)}{\partial(\psi_0, \psi_1, \psi_2, \psi_3)'} \\
&= \left(\frac{\partial\left(\xi_i^{(jj')}(t_k)\right)}{\partial \psi_0}, \frac{\partial\left(\xi_i^{(jj')}(t_k)\right)}{\partial \psi_1}, \frac{\partial\left(\xi_i^{(jj')}(t_k)\right)}{\partial \psi_2}, \frac{\partial\left(\xi_i^{(jj')}(t_k)\right)}{\partial \psi_3}\right)'.
\end{aligned} \tag{B.5}$$

First order partial derivative of $\xi_i^{(jj')}(t_k)$ with respect to $\psi_h,\ h = 0, 1, 2,$ and 3, can be derived as:

$$\begin{aligned}
\frac{\partial\left(\xi_i^{(jj')}(t_k)\right)}{\partial \psi_h} &= \frac{\partial}{\partial \psi_h}\left(\sum_{l,l'=1,2} \delta_{il}^{(j)}(t_k)\ \delta_{il'}^{(j')}(t_k) \xi_{i;ll'}^{(jj')}(t_k)\right) \\
&= \sum_{l,l'=1,2} \delta_{il}^{(j)}(t_k)\ \delta_{il'}^{(j')}(t_k) \frac{\partial}{\partial \psi_h}\left(\xi_{i;ll'}^{(jj')}(t_k)\right).
\end{aligned} \tag{B.6}$$

We can re-write the equation (B.3) as follows:

$$\begin{aligned}
\log\left(\frac{\xi_{i;ll'}^{(jj')}(t_k)}{1-\xi_{i;ll'}^{(jj')}(t_k)}\right) &= \log\left(\gamma_{i,ll'}^{(jj')}(t_k)\times N_i^{(j')}(t_k)\right) + \log\left(\frac{\pi_{il}^{(j)}(t_k)-\pi_{i;ll'}^{(jj')}(t_k)}{1-\pi_{il}^{(j)}(t_k)-\pi_{il'}^{(j')}(t_k)+\pi_{i;ll'}^{(jj')}(t_k)}\right) \\
&= \log\left(\left(\gamma_{i,ll'}^{(jj')}(t_k)\right)^{N_i^{(j')}(t_k)}\times\left(\frac{\pi_{il}^{(j)}(t_k)-\pi_{i;ll'}^{(jj')}(t_k)}{1-\pi_{il}^{(j)}(t_k)-\pi_{il'}^{(j')}(t_k)+\pi_{i;ll'}^{(jj')}(t_k)}\right)\right) \\
&= \log\left(\left(\gamma_{i,ll'}^{(jj')}(t_k)\right)^{N_i^{(j')}(t_k)}\times \mathrm{P}_{i;ll'}^{(jj')}(t_k)\right), \qquad \text{(B.7)}
\end{aligned}$$

where, $\mathrm{P}_{i;ll'}^{(jj')}(t_k) = \dfrac{\pi_{il}^{(j)}(t_k)-\pi_{i;ll'}^{(jj')}(t_k)}{1-\pi_{il}^{(j)}(t_k)-\pi_{il'}^{(j')}(t_k)+\pi_{i;ll'}^{(jj')}(t_k)}$.

Equation (B.7) can be expressed as

$$\frac{\xi_{i;ll'}^{(jj')}(t_k)}{1-\xi_{i;ll'}^{(jj')}(t_k)} = \left(\left(\gamma_{i,ll'}^{(jj')}(t_k)\right)^{N_i^{(j')}(t_k)}\times \mathrm{P}_{i;ll'}^{(jj')}(t_k)\right). \qquad \text{(B.8)}$$

We can re-write (B.8) as

$$\xi_{i;ll'}^{(jj')}(t_k) = \frac{\mathrm{P}_{i;ll'}^{(jj')}(t_k)\left(\gamma_{i;ll'}^{(jj')}(t_k)\right)^{N_i^{(j')}(t_k)}}{1+\mathrm{P}_{i;ll'}^{(jj')}(t_k)\left(\gamma_{i;ll'}^{(jj')}(t_k)\right)^{N_i^{(j')}(t_k)}}. \qquad \text{(B.9)}$$

The first order derivative of $\xi^{(jj')}_{i;ll'}(t_k)$ with respect to $\psi_h,\ h=0,1,2,$ and 3, can be derived as:

$$
\begin{aligned}
&\frac{\partial}{\partial\psi_h}\left(\xi^{(jj')}_{i;ll'}(t_k)\right)\\
&=\frac{\partial}{\partial\psi_h}\left(\frac{\mathrm{P}^{(jj')}_{i;ll'}(t_k)\left(\gamma^{(jj')}_{i;ll'}(t_k)\right)^{N_i^{(j')}}}{1+\mathrm{P}^{(jj')}_{i;ll'}(t_k)\left(\gamma^{(jj')}_{i;ll'}(t_k)\right)^{N_i^{(j')}(t_k)}}\right)\\
&=\frac{\frac{\partial}{\partial\psi_h}\left(\mathrm{P}^{(jj')}_{i;ll'}(t_k)\left(\gamma^{(jj')}_{i;ll'}(t_k)\right)^{N_i^{(j')}(t_k)}\right)\left(1+\mathrm{P}^{(jj')}_{i;ll'}(t_k)\left(\gamma^{(jj')}_{i;ll'}(t_k)\right)^{N_i^{(j')}(t_k)}-\mathrm{P}^{(jj')}_{i;ll'}(t_k)\left(\gamma^{(jj')}_{i;ll'}(t_k)\right)^{N_i^{(j')}(t_k)}\right)}{\left(1+\mathrm{P}^{(jj')}_{i;ll'}(t_k)\left(\gamma^{(jj')}_{i;ll'}(t_k)\right)^{N_i^{(j')}(t_k)}\right)^2}\\
&=\frac{\frac{\partial}{\partial\psi_h}\left(\mathrm{P}^{(jj')}_{i;ll'}(t_k)\left(\gamma^{(jj')}_{i;ll'}(t_k)\right)^{N_i^{(j')}(t_k)}\right)}{\left(1+\mathrm{P}^{(jj')}_{i;ll'}(t_k)\left(\gamma^{(jj')}_{i;ll'}(t_k)\right)^{N_i^{(j')}(t_k)}\right)^2}\\
&=\frac{\mathrm{P}^{(jj')}_{i;ll'}(t_k)\frac{\partial}{\partial\psi_h}\left(\left(\gamma^{(jj')}_{i;ll'}(t_k)\right)^{N_i^{(j')}(t_k)}\right)+\left(\gamma^{(jj')}_{i;ll'}(t_k)\right)^{N_i^{(j')}(t_k)}\frac{\partial}{\partial\psi_h}\left(\mathrm{P}^{(jj')}_{i;ll'}(t_k)\right)}{\left(1+\mathrm{P}^{(jj')}_{i;ll'}(t_k)\left(\gamma^{(jj')}_{i;ll'}(t_k)\right)^{N_i^{(j')}(t_k)}\right)^2}\\
&=\frac{\mathrm{P}^{(jj')}_{i;ll'}(t_k)N_i^{(j')}(t_k)Z_h\left(\gamma^{(jj')}_{i;ll'}(t_k)\right)^{N_i^{(j')}(t_k)}+\left(\gamma^{(jj')}_{i;ll'}(t_k)\right)^{N_i^{(j')}(t_k)}\frac{\partial}{\partial\psi_h}\left(\mathrm{P}^{(jj')}_{i;ll'}(t_k)\right)}{\left(1+\mathrm{P}^{(jj')}_{i;ll'}(t_k)\left(\gamma^{(jj')}_{i;ll'}(t_k)\right)^{N_i^{(j')}(t_k)}\right)^2}
\end{aligned}
$$

where $\gamma^{(jj')}_{i;ll'}(t_k)=e^{\boldsymbol{Z}_h\boldsymbol{\psi}_h}$.

Now,

$$\begin{aligned}
\frac{\partial}{\partial \psi_h} \mathrm{P}_{i;ll'}^{(jj')}(t_k) &= \frac{\partial}{\partial \psi_h}\left(\frac{\pi_{il}^{(j)}(t_k) - \pi_{i;ll'}^{(jj')}(t_k)}{1 - \pi_{il}^{(j)}(t_k) - \pi_{il'}^{(j')}(t_k) + \pi_{i;ll'}^{(jj')}(t_k)}\right) \\
&= \frac{\left(1 - \pi_{il}^{(j)}(t_k) - \pi_{il'}^{(j')}(t_k) + \pi_{i;ll'}^{(jj')}(t_k)\right)\left(-\frac{\partial}{\partial \psi_h}\pi_{i;ll'}^{(jj')}(t_k)\right)\left(\pi_{il}^{(j)}(t_k) - \pi_{i;ll'}^{(jj')}(t_k)\right)}{\left(1 - \pi_{il}^{(j)}(t_k) - \pi_{il'}^{(j')}(t_k) + \pi_{i;ll'}^{(jj')}(t_k)\right)^2} \\
&= \frac{-\frac{\partial}{\partial \psi_h}\pi_{i;ll'}^{(jj')}(t_k)\left(1 - \pi_{il}^{(j)}(t_k) - \pi_{il'}^{(j')}(t_k) + \pi_{i;ll'}^{(jj')}(t_k) - \pi_{il}^{(j)}(t_k) - \pi_{i;ll'}^{(jj')}(t_k)\right)}{\left(1 - \pi_{il}^{(j)}(t_k) - \pi_{il'}^{(j')}(t_k) + \pi_{i;ll'}^{(jj')}(t_k)\right)^2} \\
&= \left(\frac{1 - \pi_{il'}^{(j')}(t_k)}{\left(1 - \pi_{il}^{(j)}(t_k) - \pi_{il'}^{(j')}(t_k) + \pi_{i;ll'}^{(jj')}(t_k)\right)^2}\right)\left(-\frac{\partial}{\partial \psi_h}\pi_{i;ll'}^{(jj')}(t_k)\right).
\end{aligned} \tag{B.10}$$

The conditional joint probability, $\pi_{i;ll'}^{(jj')}(t_k)$, for the pair of events $\left(N_{il}^{(j)}(t_k), N_{il'}^{(j')}(t_k)\right)$ is defined as,

$$\pi_{i;ll'}^{(jj')}(t_k) = \begin{cases} \{a_{i;ll'}^{(jj')}(t_k) - [a_{i;ll'}^{(jj')}(t_k)^2 - 4\gamma_{i;ll'}^{(jj')}(t_k)(\gamma_{i;ll'}^{(jj')}(t_k) - 1) \times \pi_{il}^{(j)}(t_k)\pi_{il'}^{(j')}(t_k)]^{1/2}\} \times \\ \qquad \{2(\gamma_{i;ll'}^{(jj')}(t_k) - 1)\}^{-1}, \text{ if } \gamma_{i;ll'}^{(jj')}(t_k) \neq 1 \\ \pi_{il}^{(j)}(t_k)\pi_{il'}^{(j')}(t_k), \quad \text{otherwise} \end{cases}$$

where $a_{i;ll'}^{(jj')}(t_k) = 1 - (1 - \gamma_{i;ll'}^{(jj')}(t_k))(\pi_{il}^{(j)}(t_k) + \pi_{il'}^{(j')}(t_k))$.

Let, $\mathrm{A}_{i;ll'}^{(jj')}(t_k) = \pi_{il}^{(j)}(t_k) + \pi_{il'}^{(j')}(t_k)$ and $\mathrm{B}_{i;ll'}^{(jj')}(t_k) = \pi_{il}^{(j)}(t_k)\pi_{il'}^{(j')}(t_k)$.

Then, $\pi_{i;ll'}^{(jj')}(t_k)$ can be expressed as

$$\pi_{i;ll'}^{(jj')}(t_k) = \frac{\left(1 - \mathrm{A}_{i;ll'}^{(jj')}(t_k)\left(1 - \gamma_{i;ll'}^{(jj')}(t_k)\right)\right) - \left(\left(1 - \mathrm{A}_{i;ll'}^{(jj')}(t_k)\left(1 - \gamma_{i;ll'}^{(jj')}(t_k)\right)\right)^2 - 4\mathrm{B}_{i;ll'}^{(jj')}(t_k)\gamma_{i;ll'}^{(jj')}(t_k)\left(\gamma_{i;ll'}^{(jj')}(t_k) - 1\right)\right)^{1/2}}{2\left(\gamma_{i;ll'}^{(jj')}(t_k) - 1\right)}.$$

Let us define

$\mathrm{C}_{i;ll'}^{(jj')}(t_k) = 1 - \mathrm{A}_{i;ll'}^{(jj')}(t_k)\left(1 - \gamma_{i;ll'}^{(jj')}(t_k)\right)$,

$\mathrm{D}_{i;ll'}^{(jj')}(t_k) = \left(\left(1 - \mathrm{A}_{i;ll'}^{(jj')}(t_k)\left(1 - \gamma_{i;ll'}^{(jj')}(t_k)\right)\right)^2 - 4\mathrm{B}_{i;ll'}^{(jj')}(t_k)\gamma_{i;ll'}^{(jj')}(t_k)\left(\gamma_{i;ll'}^{(jj')}(t_k) - 1\right)\right)^{1/2}$, and

$\mathrm{d}_{i,ll'}^{(jj')}(t_k) = 2\left(\gamma_{i;ll'}^{(jj')}(t_k) - 1\right)\cdot$

From equation (B.10),

$$\frac{\partial}{\partial\psi_h}\pi_{i;ll'}^{(jj')}(t_k) = \frac{\partial}{\partial\psi_h}\left(\frac{\mathrm{C}_{i;ll'}^{(jj')}(t_k) - \mathrm{D}_{i;ll'}^{(jj')}(t_k)}{\mathrm{d}_{i;ll'}^{(jj')}(t_k)}\right)$$
$$= \frac{\mathrm{d}_{i;ll'}^{(jj')}(t_k)\frac{\partial}{\partial\psi_h}\left(\mathrm{C}_{i;ll'}^{(jj')}(t_k) - \mathrm{D}_{i;ll'}^{(jj')}(t_k)\right) - \left(\mathrm{C}_{i;ll'}^{(jj')}(t_k) - \mathrm{D}_{i;ll'}^{(jj')}(t_k)\right)\frac{\partial}{\partial\psi_h}\left(\mathrm{d}_{i;ll'}^{(jj')}(t_k)\right)}{\left(\mathrm{d}_{i;ll'}^{(jj')}(t_k)\right)^2} \tag{B.11}$$

where,

$$\frac{\partial}{\partial\psi_h}\left(\mathrm{C}_{i;ll'}^{(jj')}(t_k) - \mathrm{D}_{i;ll'}^{(jj')}(t_k)\right)$$
$$= \frac{\partial}{\partial\psi_h}\left(\mathrm{C}_{i;ll'}^{(jj')}(t_k)\right) - \frac{\partial}{\partial\psi_h}\left(\mathrm{D}_{i;ll'}^{(jj')}(t_k)\right)$$
$$= \frac{\partial}{\partial\psi_h}\left(1 - A_{i;ll'}^{(jj')}(t_k)\left(1-\gamma_{i;ll'}^{(jj')}(t_k)\right)\right) -$$
$$\frac{\partial}{\partial\psi_h}\left(\left(\left(1 - A_{i;ll'}^{(jj')}(t_k)\left(1-\gamma_{i;ll'}^{(jj')}(t_k)\right)\right)^2 - 4B_{i;ll'}^{(jj')}(t_k)\gamma_{i;ll'}^{(jj')}(t_k)\left(\gamma_{i;ll'}^{(jj')}(t_k)-1\right)\right)^{1/2}\right)$$
$$= A_{i;ll'}^{(jj')}(t_k)z_h\gamma_{i,ll'}^{(jj')}(t_k) - \frac{z_h\gamma_{i;ll'}^{(jj')}(t_k)\left(A_{i;ll'}^{(jj')}(t_k)C_{i;ll'}^{(jj')}(t_k) - 2B_{i;ll'}^{(jj')}(t_k)\gamma_{i;ll'}^{(jj')}(t_k) - B_{i;ll'}^{(jj')}(t_k)d_{i;ll'}^{(jj')}(t_k)\right)}{\sqrt{D_{i;ll'}^{(jj')}(t_k)}},$$

and

$$\frac{\partial}{\partial\psi_h}\left(\mathrm{d}_{i;ll'}^{(jj')}(t_k)\right) = 2z_h\gamma_{i,ll'}^{(jj')}(t_k)\cdot$$

Appendix C

R Programming Code

This chapter includes R (R Core Team, 2013) programming codes used for data manipulation and estimation of regression coefficients for marginal and association models.

```
library(boot)
library(nlme)
library(quadprog)
library(BB)
library(nleqslv)
library(numDeriv)
library(matrixcalc)
###############################################
Reading data file
###############################################
dat <- read.csv("//NLSYC.csv")
dat[1:10, ]
q<-0.80
###############################################
80th percentile for Depression of PMK
###############################################
q1.dep <- floor (quantile(dat$dep1, prob = q))
q2.dep <- floor (quantile(dat$dep2, prob = q))
q3.dep <- floor (quantile(dat$dep3, prob = q))
q4.dep <- floor (quantile(dat$dep4, prob = q))
dat$dep1.c <- ifelse (dat$dep1 >= q1.dep, 1, 0)
dat$dep2.c <- ifelse (dat$dep2 >= q2.dep, 1, 0)
dat$dep3.c <- ifelse (dat$dep3 >= q3.dep, 1, 0)
dat$dep4.c <- ifelse (dat$dep4 >= q4.dep, 1, 0)
table(dat$dep1.c)
table(dat$dep2.c)
table(dat$dep3.c)
table(dat$dep4.c)
###############################################
80th percentile for Aggression and Anxiety
###############################################
q1.agg <- floor(quantile(dat$agg1, prob = q))
q2.agg <- floor(quantile(dat$agg2, prob = q))
q3.agg <- floor(quantile(dat$agg3, prob = q))
q4.agg <- floor(quantile(dat$agg4, prob = q))
q1.anx <- floor(quantile(dat$anx1, prob = q))
```

```
q2.anx <- floor(quantile(dat$anx2, prob = q))
q3.anx <- floor(quantile(dat$anx3, prob = q))
q4.anx <- floor(quantile(dat$anx4, prob = q))
agg1 <- ifelse (dat$agg1 >= q1.agg, 1, 0)
agg2 <- ifelse (dat$agg2 >= q2.agg, 1, 0)
agg3 <- ifelse (dat$agg3 >= q3.agg, 1, 0)
agg4 <- ifelse (dat$agg4 >= q4.agg, 1, 0)
anx1 <- ifelse (dat$anx1 >= q1.anx, 1, 0)
anx2 <- ifelse (dat$anx2 >= q2.anx, 1, 0)
anx3 <- ifelse (dat$anx3 >= q3.anx, 1, 0)
anx4 <- ifelse (dat$anx4 >= q4.anx, 1, 0)
dat.anx <- cbind(anx1, anx2, anx3, anx4)
dat.agg <- cbind(agg1, agg2, agg3, agg4)
#####################################################
# Function for indicator variables delta and n.
#####################################################
delta.n <- function(dat){
K <- ncol(dat)
delta <- NULL
N <- NULL
for(i in 1:(K-1)){
del <- ifelse (dat[, i] == 1, 1, 0)
delta <- cbind(delta, del)
n <- ifelse ((dat[, i] == 1 & dat[, (i+1)] == 0) |
(dat[, i] == 0 & dat[, (i+1)] == 1), 1, 0)
N <- cbind(N, n)
}
return(list(delta = delta, N = N))
}
#####################################################
for.anx <- delta.n(dat.anx)
delta1 <- for.anx$delta
n1 <- for.anx$N # indicator variable for anxiety

for.agg <- delta.n(dat.agg)
delta2 <- for.agg$delta
n2 <- for.agg$N # indicator variable for aggression
#####################################################
sum.funct <- function(mat, p){
r <- c(mat[1 : 2])
cov.mat <- matrix(mat[3:6], ncol = 2)
d <- matrix(mat[7:length(mat)], ncol = 4*p, byrow = T)
if (is.non.singular.matrix(cov.mat) == TRUE){
inv.mat <- solve(cov.mat)
U <- t(d)%*%inv.mat%*%r
```

```
} else {
U <- rep(0, 4*p)
}
return(U = U)
}
####################################################
sum.funct1 <- function(mat){
cov.mat <- matrix(mat[3 : 6], ncol = 2)
d <- matrix(mat[7 : length(mat)], ncol = 24, byrow = T)
inv.mat <- solve(cov.mat)
M <- c(t(d)%*%inv.mat%*%d)
return(M = M)
}
####################################################
rbind.rep <- function(x, times) matrix(x, times, length(x), byrow = TRUE).
####################################################
GEE1 <- function(beta, z, delta1, delta2, n1, n2, psi){
K <- ncol(n1) + 1
p <- ncol(z)/K
# Coefficients of transition from low level of anxiety
b11 <- beta[1:p]
# Coefficients of transition from high level of anxiety
b12 <- beta[(p + 1) : (2*p)]
# Coefficients of transition from low level of aggression
b21 <- beta[(2*p + 1) : (3*p)]
# Coefficients of transition from high level of aggression
b22 <- beta[(3*p + 1) : length(beta)]
p11 <- exp(psi[1])
p22 <- exp(psi[1] + psi[4])
p12 <- exp(psi[1] + psi[2])
p21 <- exp(psi[1] + psi[3])
pp <- 2*p
dummy <- matrix(0, nrow(z), pp)
ss1 <- rep(0, 4*p)
for (i in 1 : (K-1)){
x0 <- z[, (1+(i-1)*p) : (i*p)]
####################################################
# Transition probabilities and residuals.
####################################################
# transition probabilities from low level of anxiety
pi11 <- inv.logit(x0%*%b11)
# transition probabilities from high level of anxiety
pi12 <- inv.logit(x0%*%b12)
# transition probabilities from low level of aggression
pi21 <- inv.logit(x0%*%b21)
```

```
# transition probabilities from high level of aggression
pi22 <- inv.logit(x0%*%b22)
# Conditional expectations for anxiety
mu1 <- delta1[, i]*pi11 + (1 - delta1[, i])*pi12
# Conditional expectations for aggression
mu2 <- delta2[, i]*pi21 + (1 - delta2[, i])*pi22
# Error terms
r <- cbind(n1[, i] - mu1, n2[, i] - mu2)
####################################################
# Working covariance matrix.
####################################################
a11 <- 1 - (1 - p11)*(pi11 + pi21)
a22 <- 1 - (1 - p22)*(pi12 + pi22)
a12 <- 1 - (1 - p12)*(pi11 + pi22)
a21 <- 1 - (1 - p21)*(pi12 + pi21)
####################################################
# Conditional joint probabilities.
####################################################
qi11 <- ifelse (p11 != 1, ((a11 - sqrt(a11^2 - 4*p11*(p11 - 1)*pi11*pi21))/
(2*(p11 - 1))), (pi11*pi21))
qi22 <- ifelse (p22 != 1, ((a22 - sqrt(a22^2 - 4*p22*(p22 - 1)*pi12*pi22))/
(2*(p22 - 1))), (pi12*pi22))
qi12 <- ifelse (p12 != 1, ((a12 - sqrt(a12^2 - 4*p12*(p12 - 1)*pi11*pi22))/
(2*(p12 - 1))), (pi11*pi22))
qi21 <- ifelse (p21 != 1, ((a21 - sqrt(a21^2 - 4*p21*(p21 - 1)*pi12*pi21))/
(2*(p21 - 1))), (pi12*pi21))
w1 <- delta1[, i]*delta2[, i]*(qi11 - pi11*pi21)
w2 <- (1-delta1[, i])*(1 - delta2[, i])*(qi22 - pi12*pi22)
w3 <- delta1[, i]*(1 - delta2[, i])*(qi12 - pi11*pi22)
w4 <- (1-delta1[, i])*delta2[, i]*(qi21 - pi12*pi21)
v12 <- w1 + w2 + w3 + w4
v11 <- delta1[, i]*pi11*(1 - pi11) + (1 - delta1[, i])*pi12*(1 - pi12)
v22 <- delta2[, i]*pi21*(1 - pi21) + (1 - delta2[, i])*pi22*(1 - pi22)
cov.mat <- cbind(v11, v12, v12, v22) # covariance matrix
##################################################################
# Partial derivatives of mean w.r.t. marginal transition parameters.
##################################################################
u1 <- delta1[, i]*c(pi11*(1 - pi11))*x0
u2 <- (1 - delta1[, i])*c(pi12*(1 - pi12))*x0
u3 <- delta2[, i]*c(pi21*(1 - pi21))*x0
u4 <- (1 - delta2[, i])*c(pi22*(1 - pi22))*x0
u  <- cbind(u1, u2, dummy, dummy, u3, u4)
###########################################################
# Estimating equations for marginal transition parameters.
###########################################################
```

```
big.mat <- cbind(r, cov.mat, u)
ss1 <- ss1 + rowSums(apply(big.mat, 1, sum.funct, p = p))
}
return(ss1)
}
###################################################################################
# We use single covariate depression of PMK to estimate marginal transition
# parameters for anxiety and aggression.
###################################################################################
z <- cbind(1, dat$dep1.c, 1, dat$dep2.c, 1, dat$dep3.c, 1, dat$dep4.c)
b11 <- rep(0, 2) # initial values for marginal transition parameters
b12 <- rep(0, 2) # initial values for marginal transition parameters
b21 <- rep(0, 2) # initial values for marginal transition parameters
b22 <- rep(0, 2) # initial values for marginal transition parameters
beta <- c(b11, b12, b21, b22)
psi <- rep(1, 4) # initial values for association parameters
###################################################################################
# R function "nleqslv" is used to solve systems of nonlinear equations to estimate
# the marginal transition parameters.
###################################################################################
est <- nleqslv(beta, GEE1, z = z, delta1 = delta1, delta2 = delta2,
       n1 = n1, n2 = n2, psi = psi, global = "dbldog",
       control = list(maxit = 1500, xtol = 1e-5, ftol = 1e-5, trace = 1))
###################################################################################
# Number of siblings is added in this model to estimate its effect
# on marginal transition probabilities for anxiety and aggression.
###################################################################################
z <- cbind(1, dat$dep1.c, dat$sib1.c,
           1, dat$dep2.c, dat$sib2.c,
           1, dat$dep3.c, dat$sib3.c,
           1, dat$dep4.c, dat$sib4.c)
# estimates obtained from previous model is used as initial values for this model.
b11 <- c(est$x[1 : 2], 0)
b12 <- c(est$x[3 : 4], 0)
b21 <- c(est$x[5 : 6], 0)
b22 <- c(est$x[7 : 8], 0)
beta <- c(b11,b12,b21,b22)
psi <- rep(1,4)
est1 <- nleqslv(beta, GEE1, z = z, delta1 = delta1, delta2 = delta2,
                n1 = n1, n2 = n2, psi = psi, global = "dbldog",
                control = list(maxit = 1500, xtol = 1e-5, ftol = 1e-5, trace = 1))
###################################################################################
# Gender of a child is added to see its effects alongwith previous two covariates
# on marginal transition probabilities.
###################################################################################
```

```
z <- cbind(1, dat$dep1.c, dat$gender.f, dat$sib1.c,
           1, dat$dep2.c, dat$gender.f, dat$sib2.c,
           1, dat$dep3.c, dat$gender.f, dat$sib3.c,
           1, dat$dep4.c, dat$gender.f, dat$sib4.c)
b11 <- c(est1$x[1 : 3], 0)
b12 <- c(est1$x[4 : 6], 0)
b21 <- c(est1$x[7 : 9], 0)
b22 <- c(est1$x[10 : 12], 0)
beta <- c(b11, b12, b21, b22)
psi <- rep(1,4)
est2 <- nleqslv(beta, GEE1, z = z, delta1 = delta1, delta2 = elta2,
       n1 = n1, n2 = n2, psi = psi, global = "dbldog",
       control = list(maxit = 1500, xtol = 1e-5, ftol = 1e-5, trace = 1))
####################################################################################
# Finally family status if the child is added in the model.
####################################################################################
z <- cbind(1, dat$dep1.c, dat$gender.f, dat$sib1.c, dat$pstat1.c,
           1, dat$dep2.c, dat$gender.f, dat$sib2.c, dat$pstat2.c,
           1, dat$dep3.c, dat$gender.f, dat$sib3.c, dat$pstat3.c,
           1, dat$dep4.c, dat$gender.f, dat$sib4.c, dat$pstat4.c)
b11 <- c(est2$x[1:4],0)
b12 <- c(est2$x[5:8],0)
b21 <- c(est2$x[9:12],0)
b22 <- c(est2$x[13:16],0)
beta <- c(b11,b12,b21,b22)
psi <- rep(1,4)
est3 <- nleqslv(beta, GEE1, z = z, delta1 = delta1, delta2 = delta2,
        n1 = n1, n2 = n2, psi = psi, global = "pwldog",
        control = list(maxit = 1500, xtol = 1e-5, ftol = 1e-5, trace = 1))
####################################################
zi <- function(psi, pidn){
pi11 <- pidn[1]
pi12 <- pidn[2]
pi21 <- pidn[3]
pi22 <- pidn[4]
d11 <- pidn[5]
d12 <- pidn[6]
d21 <- pidn[7]
d22 <- pidn[8]
n2 <- pidn[9]
p11 <- exp(psi[1])
p22 <- exp(psi[1] + psi[4])
p12 <- exp(psi[1] + psi[2])
p21 <- exp(psi[1] + psi[3])
a11 <- 1 - (1 - p11)*(pi11 + pi21)
```

```
a22 <- 1 - (1 - p22)*(pi12 + pi22)
a12 <- 1 - (1 - p12)*(pi11 + pi22)
a21 <- 1 - (1 - p21)*(pi12 + pi21)

qi11 <- ifelse (p11 != 1, ((a11 - sqrt(a11^2 - 4*p11*(p11 - 1)*pi11*pi21))/
               (2*(p11 - 1))), (pi11*pi21))
qi22 <- ifelse (p22 != 1, ((a22 - sqrt(a22^2 - 4*p22*(p22 - 1)*pi12*pi22))/
               (2*(p22 - 1))), (pi12*pi22))
qi12 <- ifelse (p12 != 1, ((a12 - sqrt(a12^2 - 4*p12*(p12 - 1)*pi11*pi22))/
               (2*(p12 - 1))), (pi11*pi22))
qi21 <- ifelse (p21 != 1, ((a21 - sqrt(a21^2 - 4*p21*(p21 - 1)*pi12*pi21))/
               (2*(p21 - 1))), (pi12*pi21))
ri11 <- (pi11 - qi11)/(1 - pi11 - pi21 + qi11)
ri12 <- (pi11 - qi12)/(1 - pi11 - pi22 + qi12)
ri21 <- (pi12 - qi21)/(1 - pi12 - pi21 + qi21)
ri22 <- (pi12 - qi22)/(1 - pi12 - pi22 + qi22)
zi11 <- (ri11*(p11^n2))/(1 + ri11*(p11^n2))
zi22 <- (ri22*(p22^n2))/(1 + ri22*(p22^n2))
zi12 <- (ri12*(p12^n2))/(1 + ri12*(p12^n2))
zi21 <- (ri21*(p21^n2))/(1 + ri21*(p21^n2))
return(d11*zi11 + d12*zi12 + d21*zi21 + d22*zi22)
}
####################################################
grad.zi <- function(psi, pidn){
return(grad(zi, psi, method = "simple", pidn = pidn))
}
####################################################
zi.beta <- function(beta, dx){
p <- dx[5]
b11 <- beta[1:p]
b12 <- beta[(p + 1) : (2*p)]
b21 <- beta[(2*p + 1) : (3*p)]
b22 <- beta[(3*p + 1) : length(beta)]
p11 <- exp(dx[1])
p22 <- exp(dx[1] + dx[4])
p12 <- exp(dx[1] + dx[2])
p21 <- exp(dx[1] + dx[3])
d11 <- dx[6]
d12 <- dx[7]
d21 <- dx[8]
d22 <- dx[9]
n2 <- dx[10]
x0 <- dx[11:length(dx)]
pi11 <- inv.logit(t(x0)%*%b11)
pi12 <- inv.logit(t(x0)%*%b12)
```

```
pi21 <- inv.logit(t(x0)%*%b21)
pi22 <- inv.logit(t(x0)%*%b22)
a11 <- 1 - (1 - p11)*(pi11 + pi21)
a22 <- 1 - (1 - p22)*(pi12 + pi22)
a12 <- 1 - (1 - p12)*(pi11 + pi22)
a21 <- 1 - (1 - p21)*(pi12 + pi21)
qi11 <- ifelse (p11 != 1, ((a11 - sqrt(a11^2 - 4*p11*(p11 - 1)*pi11*pi21))/
               (2*(p11 - 1))), (pi11*pi21))
qi22 <- ifelse (p22 != 1, ((a22 - sqrt(a22^2 - 4*p22*(p22 - 1)*pi12*pi22))/
              (2*(p22 - 1))), (pi12*pi22))
qi12 <- ifelse (p12 != 1, ((a12 - sqrt(a12^2 - 4*p12*(p12 - 1)*pi11*pi22))/
              (2*(p12 - 1))), (pi11*pi22))
qi21 <- ifelse (p12 != 1, ((a21 - sqrt(a21^2 - 4*p21*(p21 - 1)*pi12*pi21))/
              (2*(p21 - 1))), (pi12*pi21))
ri11 <- (pi11 - qi11)/(1 - pi11 - pi21 + qi11)
ri12 <- (pi11 - qi12)/(1 - pi11 - pi22 + qi12)
ri21 <- (pi12 - qi21)/(1 - pi12 - pi21 + qi21)
ri22 <- (pi12 - qi22)/(1 - pi12 - pi22 + qi22)
zi11 <- (ri11*(p11^n2))/(1 + ri11*(p11^n2))
zi22 <- (ri22*(p22^n2))/(1 + ri22*(p22^n2))
zi12 <- (ri12*(p12^n2))/(1 + ri12*(p12^n2))
zi21 <- (ri21*(p21^n2))/(1 + ri21*(p21^n2))
return(d11*zi11 + d12*zi12 + d21*zi21 + d22*zi22)
}
#####################################################
grad.zi.beta <- function(beta, dx){
return(grad(zi.beta, beta, method = "simple", dx = dx))
}
#####################################################
GEE2 <- function(psi, z, delta1, delta2, n1, n2, beta){
K <- ncol(n1) + 1
p <- ncol(z)/K
b11 <- beta[1 : p]
b12 <- beta[(p+1) : (2*p)]
b21 <- beta[(2*p+1) : (3*p)]
b22 <- beta[(3*p+1) : length(beta)]
p11 <- exp(psi[1])
p22 <- exp(psi[1] + psi[4])
p12 <- exp(psi[1] + psi[2])
p21 <- exp(psi[1] + psi[3])
rr <- rep(0, 4)
for(i in 1 : (K-1)){
d11 <- ifelse ((delta1[, i] == 1 &  delta2[, i] == 1), 1, 0)
d12 <- ifelse ((delta1[, i] == 1 &  delta2[, i] == 0), 1, 0)
d21 <- ifelse ((delta1[, i] == 0 &  delta2[, i] == 1), 1, 0)
```

```
d22 <- ifelse ((delta1[, i] == 0 &  delta2[, i] == 0), 1, 0)
x0 <- z[, (1 + (i - 1)*p) : (i*p)]
####################################################
pi11 <- inv.logit(x0%*%b11)
pi12 <- inv.logit(x0%*%b12)
pi21 <- inv.logit(x0%*%b21)
pi22 <- inv.logit(x0%*%b22)
a11 <- 1 - (1 - p11)*(pi11 + pi21)
a22 <- 1 - (1 - p22)*(pi12 + pi22)
a12 <- 1 - (1 - p12)*(pi11 + pi22)
a21 <- 1 - (1 - p21)*(pi12 + pi21)
qi11 <- ifelse (p11 != 1, ((a11 - sqrt(a11^2 - 4*p11*(p11 - 1)*pi11*pi21))/
              (2*(p11 - 1))), (pi11*pi21))
qi22 <- ifelse (p22 != 1,((a22 - sqrt(a22^2 - 4*p22*(p22 - 1)*pi12*pi22))/
              (2*(p22 - 1))), (pi12*pi22))
qi12 <- ifelse (p12 != 1, ((a12 - sqrt(a12^2 - 4*p12*(p12 - 1)*pi11*pi22))/
              (2*(p12 - 1))), (pi11*pi22))
qi21 <- ifelse (p12 != 1,((a21 - sqrt(a21^2 - 4*p21*(p21 - 1)*pi12*pi21))/
              (2*(p21 - 1))), (pi12*pi21))
zi11 <- (qi11*(p11^n2[, i]))/(1+qi11*(p11^n2[, i]))
ri11 <- (pi11 - qi11)/(1 - pi11 - pi21 + qi11)
ri12 <- (pi11 - qi12)/(1 - pi11 - pi22 + qi12)
ri21 <- (pi12 - qi21)/(1 - pi12 - pi21 + qi21)
ri22 <- (pi12 - qi22)/(1 - pi12 - pi22 + qi22)
zi11 <- (ri11*(p11^n2[, i]))/(1 + ri11*(p11^n2[, i]))
zi22 <- (ri22*(p22^n2[, i]))/(1 + ri22*(p22^n2[, i]))
zi12 <- (ri12*(p12^n2[, i]))/(1 + ri12*(p12^n2[, i]))
zi21 <- (ri21*(p21^n2[, i]))/(1 + ri21*(p21^n2[, i]))
zi0 <- d11*zi11 + d12*zi12 + d21*zi21 + d22*zi22
eps <- n1[, i] - zi0
ss0 <- zi0*(1 - zi0)
pidn <- cbind(pi11, pi12, pi21, pi22, d11, d12, d21, d22, n2[, i])
ss <- apply(pidn, 1, grad.zi, psi = psi)
ss1 <- ss%*%(eps/ss0)
rr <- rr + ss1
}
return(rr)
}
####################################################
psi < -rep (1, 4)
beta <- est3$x
est.psi <- nleqslv(psi, GEE2, z = z, delta1 = delta1, delta2 = delta2,
                  n1 = n1, n2 = n2, beta = beta, control = list(maxit = 500))
psi <- est.psi$x
GEE2(psi, z = z, delta1 = delta1, delta2 = delta2, n1 = n1, n2 = n2, beta = beta)
```

```
####################################################
final.estimation <- function(beta, z = z, delta1 = delta1, delta2 = delta2,
n1 = n1, n2 = n2, psi = psi, tolbeta, tolpsi, tol1, tol2, iter){
beta0 <- beta
psi0 <- psi
i <- 0
repeat{
i <- i + 1
ss <- nleqslv(beta0, GEE1, z = z, delta1 = delta1, delta2 = delta2,
              n1 = n1, n2 = n2, psi = psi0,
              control = list(xtol = tolbeta, ftol = tolbeta, maxit = 500))
beta <- ss$x
print(beta)
ss1 <- nleqslv(psi0, GEE2, z = z, delta1 = delta1, delta2 = delta2,
               n1 = n1, n2 = n2, beta = beta,
               control = list(xtol = tolpsi, ftol = tolpsi, maxit = 500))
psi <- ss1$x
print(psi)
print(i)
cc1 <- abs(beta0 - beta)
c1 <- max(cc1)
c10 <- which(cc1 == c1)
cc2 <- abs(psi0 - psi)
c2 <- max(cc2)
c20 <- which(cc2 == c2)
print(c(c1, c2))
print(c(c10, c20))
if (c1 < tol1 && c2 < tol2){
print("convergence achieved")
ss <- nleqslv(beta0, GEE1, z = z, delta1 = delta1, delta2 = delta2,
              n1 = n1, n2 = n2, psi = psi0, jacobian = TRUE,
              control = list(xtol = tol1, ftol = tol2, maxit = 500))
ss1 <- nleqslv(psi0, GEE2, z = z, delta1 = delta1, delta2 = delta2,
               n1 = n1, n2 = n2, beta = beta, jacobian = TRUE,
               control = list(xtol = tol1, ftol = tol2, maxit = 500))
break
}
else{
beta0 <- beta
psi0 <- psi
}
if (i > iter){
print("iter limit exceeds without convergence")
break
}
```

```
}
return(list(beta.est = ss, psi.est = ss1))
}
####################################################
z <- cbind(1, dat$dep1.c, dat$gender.f, dat$sib1.c, dat$pstat1.c,
           1, dat$dep2.c, dat$gender.f, dat$sib2.c, dat$pstat2.c,
           1, dat$dep3.c, dat$gender.f, dat$sib3.c, dat$pstat3.c,
           1, dat$dep4.c, dat$gender.f, dat$sib4.c, dat$pstat4.c)
beta <- est3$x
psi <- rep(1, 4)
final.est <- final.estimation(beta, z, delta1, delta2, n1, n2, psi,
                              1e-5, 1e-5, 1e-5, 1e-5, 100)
beta <- final.est$beta.est$x
ps i<- final.est$psi.est$x
jac.beta <- final.est$beta.est$jac
jac.psi <- final.est$psi.est$jac
####################################################
U <- function(x){
x1 <- as.vector(x)
return(c(x1%*%t(x1)))
}
####################################################
GEE <- function(beta, z, delta1, delta2, n1, n2, psi, jac.beta, jac.psi){
K <- ncol(n1) + 1
p <- ncol(z)/K
b11 <- beta[1 : p]
b12 <- beta[(p + 1) : (2*p)]
b21 <- beta[(2*p + 1) : (3*p)]
b22 <- beta[(3*p + 1) : length(beta)]
p11 <- exp(psi[1])
p22 <- exp(psi[1] + psi[4])
p12 <- exp(psi[1] + psi[2])
p21 < -exp(psi[1] + psi[3])
pp <- 2*p
dummy <- matrix(0, nrow(z), pp)
fg <- matrix(0, (length(beta) + length(psi)), (length(beta) + length(psi)))
fg1 <- matrix(0, length(psi), length(beta))
for(i in 1 : (K-1)){
x0 <- z[, (1+(i-1)*p) : (i*p)]
####################################################
# calculation of pi and R
####################################################
pi11 <- inv.logit(x0%*%b11)
pi12 <- inv.logit(x0%*%b12)
pi21 < -inv.logit(x0%*%b21)
```

```
pi22 <- inv.logit(x0%*%b22)
mu1 <- delta1[, i]*pi11 + (1 - delta1[, i])*pi12
mu2 <- delta2[, i]*pi21 + (1 - delta2[, i])*pi22
r <- cbind(n1[, i] - mu1, n2[, i] - mu2)
#####################################################
# calculation of V
#####################################################
a11 <- 1 - (1 - p11)*(pi11 + pi21)
a22 <- 1 - (1 - p22)*(pi12 + pi22)
a12 <- 1 - (1 - p12)*(pi11 + pi22)
a21 <- 1 - (1 - p21)*(pi12 + pi21)
qi11 <- ifelse (p11 != 1, ((a11 - sqrt(a11^2 - 4*p11*(p11 - 1)*pi11*pi21))/
               (2*(p11 - 1))), (pi11*pi21))
qi22 <- ifelse (p22 != 1,((a22 - sqrt(a22^2 - 4*p22*(p22 - 1)*pi12*pi22))/
               (2*(p22 - 1))), (pi12*pi22))
qi12 <- ifelse (p12 != 1,((a12 - sqrt(a12^2 - 4*p12*(p12 - 1)*pi11*pi22))/
               (2*(p12 - 1))), (pi11*pi22))
qi21 <- ifelse (p21 != 1,((a21-sqrt(a21^2 - 4*p21*(p21 - 1)*pi12*pi21))/
               (2*(p21 - 1))), (pi12*pi21))
w1 <- delta1[, i]*delta2[, i]*(qi11 - pi11*pi21)
w2 <- (1 - delta1[, i])*(1 - delta2[, i])*(qi22 - pi12*pi22)
w3 <- delta1[, i]*(1 - delta2[, i])*(qi12 - pi11*pi22)
w4 <- (1 - delta1[, i])*delta2[, i]*(qi21 - pi12*pi21)
v12 <- w1 + w2 + w3 + w4
v11 <- delta1[, i]*pi11*(1 - pi11) + (1 - delta1[, i])*pi12*(1 - pi12)
v22 <- delta2[, i]*pi21*(1 - pi21) + (1 - delta2[, i])*pi22*(1 - pi22)
cov.mat <- cbind(v11, v12, v12, v22)
#####################################################
# Calculation of D
#####################################################
u1 <- delta1[, i]*c(pi11*(1 - pi11))*x0
u2 <- (1 - delta1[, i])*c(pi12*(1 - pi12))*x0
u3 <- delta2[, i]*c(pi21*(1 - pi21))*x0
u4 <- (1 - delta2[, i])*c(pi22*(1 - pi22))*x0
u <- cbind(u1, u2, dummy, dummy, u3, u4)
#########################################################
big.mat <- cbind(r, cov.mat, u)
ss1 <- t(apply(big.mat, 1, sum.funct, p=p))
#########################################################
d11 <- ifelse ((delta1[, i] == 1 &  delta2[, i] == 1), 1, 0)
d12 <- ifelse ((delta1[, i] == 1 &  delta2[, i] == 0), 1, 0)
d21 <- ifelse ((delta1[, i] == 0 &  delta2[, i] == 1), 1, 0)
d22 <- ifelse ((delta1[, i] == 0 &  delta2[, i] == 0), 1 ,0)
ri11 <- (pi11 - qi11)/(1 - pi11 - pi21 + qi11)
ri12 <- (pi11 - qi12)/(1 - pi11 - pi22 + qi12)
```

```
ri21 <- (pi12 - qi21)/(1 - pi12 - pi21 + qi21)
ri22 <- (pi12 - qi22)/(1 - pi12 - pi22 + qi22)
zi11 <- (ri11*(p11^n2[, i]))/(1 + ri11*(p11^n2[, i]))
zi22 <- (ri22*(p22^n2[, i]))/(1 + ri22*(p22^n2[, i]))
zi12 <- (ri12*(p12^n2[, i]))/(1 + ri12*(p12^n2[, i]))
zi21 <- (ri21*(p21^n2[, i]))/(1 + ri21*(p21^n2[, i]))
zi0 <- d11*zi11 + d12*zi12 + d21*zi21 + d22*zi22
eps <- as.vector(n1[, i] - zi0)
ss0 <- as.vector(zi0*(1 - zi0))
pidn <- cbind(pi11, pi12, pi21, pi22, d11, d12, d21, d22, n2[, i])
tt <- apply(pidn, 1, grad.zi, psi = psi)
tt1 <- t(tt)*(eps/ss0)
#####################################################
st <- cbind(ss1, tt1)
st1 <- matrix(rowSums(apply(st, 1, U)), ncol = (length(beta) + length(psi)))
fg <- fg + st1
#####################################################
psip <- c(psi, p)
dx <- cbind(rbind.rep(psip, nrow(z)), d11, d12, d21, d22, n2[, i], x0)
hj <- apply(dx, 1, grad.zi.beta, beta = beta)
hj1 <- matrix(0, length(psi), length(beta))
for(j in 1 : nrow(z)){
hj1 <- hj1 + (t(tt)[j, ]%*%t(hj[, j]))/ss0[j]
}
fg1 <- fg1 + hj1
}
hh0 <- fg/((K - 1)*nrow(z))
hh <- cbind(rbind(jac.beta, fg1), rbind(matrix(0, length(beta), length(psi)),
jac.psi))/((K - 1)*nrow(z))
hh.inv <- solve(hh)
#####################################################
# Variance-covariance matrix
#####################################################
se <- sqrt(diag((hh.inv%*%hh0%*%t(hh.inv))/((K - 1)*nrow(z))))
#####################################################
test <- c(beta, psi)/se
pval <- round(2*pnorm( - abs(test)), digits = 3)
process1.01 <- cbind(coef = beta[1 : p],
                     OR = exp(beta[1 : p]),
                     se = se[1 : p], pvalue = pval[1 : p])
process1.10 <- cbind(coef = beta[(p + 1) : (2*p)],
                     OR = exp(beta[(p + 1) : (2*p)]),
                     se = se[(p + 1) : (2*p)],
                 pvalue = pval[(p + 1) :(2*p)])
process2.01 <- cbind(coef = beta[(2*p + 1) : (3*p)],
```

```
                       OR = exp(beta[(2*p + 1) : (3*p)]),
                       se = se[(2*p + 1) : (3*p)],
                   pvalue = pval[(2*p + 1) : (3*p)])
process2.10 <- cbind(coef = beta[(3*p + 1) : length(beta)],
                       OR = exp(beta[(3*p + 1) : length(beta)]),
                       se = se[(3*p + 1) : length(beta)],
                   pvalue = pval[(3*p + 1) : length(beta)])

psi.summary <- cbind(coef = psi,
                       OR = exp(psi),
                       se = se[(length(beta) + 1) : length(pval)],
                       pval[(length(beta) + 1) : length(pval)])
####################################################
# Estimates of OR
####################################################
process.OR <- matrix(exp(c(psi[1],
                           psi[1] + psi[4],
                           psi[1] + psi[2],
                           psi[1] + psi[3])), nrow = 1)
colnames(process.OR) <- c("tran11", "tran22", "tran12", "tran21")

results <- list(process1.01 = process1.01,
                process1.10 = process1.10,
                process2.01 = process2.01,
                process2.10 = process2.10,
                psi.summary = psi.summary,
                 process.OR = process.OR)
results
}
####################################################
# Final Estimates
####################################################
final.results <- GEE(beta, z, delta1, delta2, n1, n2, psi, jac.beta, jac.psi)
final.results
```